GOODMAN, JOSHUA, Ph.D. An Examination of the Residual Covariance Structures of Complex Performance Assessments Under Various Scaling and Scoring Methods. (2008) Directed by Dr. Richard M. Luecht. 190 pp.

Large-scale assessment programs are increasingly including complex performance exercises along with traditional multiple-choice items in a given test. These performance assessments are developed in part to measure sets of skills that are part of the trait to be measured, but are not easily assessed with multiple choice items. One approach to creating and scoring these items is to create a set of tasks within a scenario that can be objectively scored using a set of scoring rules to yield dichotomous responses. Including complex performance items introduces two potential challenges: first, the performance items are developed to measure something distinctly different and may introduce some degree of multidimensionality into the test; second, as the set of measurement opportunities stem from a common stimuli and are scored with a set of elaborate rules, contextual and scoring dependencies are likely to arise. Both multidimensionality and statistical dependencies may create a situation where non-zero residual covariances are present. This study uses a computer simulation to create different amounts of association among the CPE item due to the three sources mentioned above. The magnitude and distribution of the residual covariances are assessed under two different methods for scoring the simulations (dichotomous or polytomous scoring) and under different Item Response Theory based scaling methods (creating separate scales for the two item types or joint calibrations of all items).

The results indicate the following: If only contextual/scoring dependencies are present in the data, polytomous scoring is effective in eliminating some of the extreme

dependencies due to scoring factors, but does not decrease the average amount of residual covariance among the measurement opportunities of the performance items. Treating performance exercises and selected response items as two separate and distinct scales was effective in controlling the amount of residual covariance regardless of the underlying dimensional structure. However, when the correlation between traits was moderate to high, the joint calibration approaches show similar amounts of residual covariance among performance exercises as the separate scale approach, and produce score estimates that are more precise. Last, when dependencies are the result of all the sources mentioned above, only the separate scales approach couple with the polytomous scoring approach is successful in reducing the residual covariance to zero levels. Choosing a joint scaling approach and polytomously scored items when the data is two-dimensional, even when context or scoring dependencies are present, leads to large amounts of residual covariance.

AN EXAMINATION OF THE RESIDUAL COVARIANCE STRUCTURES OF COMPLEX PERFORMANCE EXERCISES UNDER VARIOUS SCALING AND SCORING METHODS

By

Joshua T Goodman

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2008

Approved by

_____
Committee Chair

For Janie, Sam, and a player yet to be named.

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

For the better part of a century educational institutions, credentialing associations, and certification boards in the United States of America have relied, at least in part, on large-scale assessments to render high-stakes decisions about which students are promoted or accepted to colleges and which professionals are allowed to practice their chosen occupations. In the early 20th century, high-stakes examinations typically consisted of written tasks or presentations that were scored by expert raters (2000). The first edition of Educational Measurement (Lindquest, 1951) contained a chapter dedicated to scores produced by expert raters. However, even by this time, widespread skepticism surrounded the reliability of the scores in these examples of performance assessment (PA) in the emerging measurement community (Hartog & Rhodes, 1936; Stalnaker, 1951). By the time of Coffman's (1971) chapter in the third edition of Educational Measurement, the use of essay test and other forms of PA had fallen largely out of favor in the testing community. By the printing of the third edition of Educational Measurement (Linn, 1989), performance assessment did not warrant a single chapter (Clauser, 2000).

Coupled with this decline was the rise of the selected-response or multiple choice examinations. Multiple-choice tests, which offered the benefits of objective scoring and the inclusion of a great number of items covering a wide breadth of content, soon became

the dominant format of large-scale assessments. The development of modern test theory and new measurement models only strengthened the appeal of the highly reliable multiple-choice tests in the measurement community, particularly in standardized testing programs with a high volume of test-takers.

This trend began to change in the mid 1980's, which saw a reincarnation of PAs in a much modified form. For the purpose of this study, PA will be broadly be defined as the processes in which test-takers demonstrate their ability to apply knowledge and skills and/or to put knowledge and understanding into action in tasks designed to simulate real-life situations (Muraki, Hombo, & Lee, 2000; Wiggins, 1993). The newly reborn PAs have again begun to appear along-side multiple-choice items in large-scale assessment. A number of sources motivate this reemergence. First, a wave of educational reform, emphasizing higher order thinking skills (e.g. problem solving, reasoning, and critical thinking), surfaced in the United States in the 1980's (Hambleton, 2000; Resnick & Resnick, 1992). These new objectives proved difficult to assess validly in a multiple-choice format. Second, credentialing and certification boards recognized the need to directly assess the knowledge base and/or task specific skills that professionals in practice are likely to encounter. If the test is to be a true measure of the knowledge, skills, and abilities required for on-the-job success, then both the discrete knowledge and the skills specific tasks must be measured. As in educational testing, these skills specific traits are difficult to measure in a multiple-choice format.

Performance Assessments in Modern Standardized Tests

Hambleton et al (2000) suggest that in the United States, 40 states employ some form of performance assessment in their state-sponsored testing programs. Similarly, a survey by Lane (2005) states that 63 percent of state assessments included performance assessments along with multiple choice items. Under the No Child Left Behind Act, all states are currently required to assess writing. In these writing assessments, students are given a writing prompt and are asked to produce a writing sample in response. PA in schools is not limited to writing assessments. A great number of states (e.g. California, Delaware, Kentucky, Maryland, Oregon) routinely use performance assessments in large-scale high-stakes examinations in subjects such as science, mathematics, reading, social studies, and computer competency (Ferrara & et al., 1997; Ferrara, Huynh, & Michaels, 1999; Niemi, Baker, & Sylvester, 2007; Pearson, Calfee, Webb, & Fleischer, 2002). Performance assessment in education is not limited to state-wide testing programs, but is also found in many standardized college admittance tests. The newest version Test of English as a Foreign Language (TOEFL iBT) PAs are used to assess proficiency in all four language modalities—that is, reading, writing, listening and speaking (Educational Testing Service, 2007). The latest version of the SAT includes a writing assessment (Camara, 2003) and the Advanced Placement tests have long included complex writing tasks and open-ended problems (Bennett & Wadkins, 1995).

The trend to include PAs in large-scale testing has also influenced the practices of credentialing/certification/licensure agencies. A number of credentialing and certification agencies include performance assessments as a component in examination

programs. The Architect Registration Exam (ARE) contains subsections of that require examinees to respond to open-ended questions that require drawing or other graphical representation in response to a problem or situation (Bejar, 1988; Bridgeman, Bejar, & Friedman, 1999). The United States Medical Licensing Exam (USMLE) contains several patient-case simulations where examinees are presented a medical case and must then order medical tests, make diagnoses, and watch patient progress over time (Clyman, Melnick, Clauser, Mancall, & Bashook, 1995). The American Institute of Certified Public Accountants (AICPA) makes use of simulation items in the United Certified Public Accountant Examination (UCPAE). Examinees are presented with an accounting scenario, and a series of accounting tasks (e.g. composing audit letters, assessing risk, spreadsheet applications) that are all tied to the same context (DeVore, 2004; Goodman & Luecht, 2007).

Complex Performance Assessments

The reemergence of PAs is rooted in a desire to measure things that are difficult to assess with multiple-choice questions. To meet this need, the traditional performance assessments of the past, namely writing samples and oral presentations or interviews, must be expanded to include a wide variety of new item types and new topics (Hambleton, 2000; Parshall, Davey, & Pashley, 2000). Advances in technology, specifically the widespread availability of computers with multimedia capabilities, memory capacity, and graphical interfaces, coupled with an expansion of measurement models and computing power to fit these models have facilitated the transition of

traditional PAs into a new generation of complex performance assessments. (Bennett, 1999; Bennett, Bartram, & Hambleton, 2006; Parshall et al., 2000).

Complex performance exercises (CPE) typically consist of a series of related tasks that are presented to examinees. The examinee's response (or collection of responses) is then scored. CPEs lend themselves to the traditional domain of performance type items (essays or presentation) and provide opportunities for innovative item types like sequential problem solving (Bennett, Morley, & Quardt, 2000), simulation of real-life tasks (Clyman et al., 1995; DeVore, 2004), or multimedia based items (Bennett, 2001; Bennett et al., 1999). The advantage of both including innovative CPE items and/or traditional performance assessments is that they can be carefully designed to approximate the activities or processes of interest in a direct manner. For instance, a CPE included in a computer competency test could present a test taker with a word processor document. The test taker could then be asked to complete tasks that more closely resemble what is expected of "computer literature" student—e.g. bold certain passages of text, cut and paste a specific piece of text in a different location, etc. A CPE of this nature could be considered authentic (Linn & Burton, 1994), as the task itself closely resembles what might be expected in the practice outside of the testing environment. It is no stretch of the imagination, that with careful design, CPEs could be designed to assess complicated sets of skills and processes in both educational and certification testing.

Such items are also touted as interactive (Bachman, 2002; Messick 1994; Messick, 1995; Messick & Hakel, 1998). In the example cited above, the item requires the examinee to engage the problem in the same way he/she would in practice, and

success can only be achieved if the required word processing skills are present and activated. No degree of test taking skill would ensure success, unlike a selected-response (SR) question where cuing and a general test savviness would influence correct responses.

Messick, in his seminal work on validity in the 3rd Edition of Educational Measurement (1989), states that it is the interpretation and use of scores that must be the focus in considering the validity of an instrument and that the strength of the inference between item and the construct must first be considered. Testing, by its very nature is an artificial process, as test-takers are using employing skills and knowledge to complete tasks in a context that is removed from the actual settings where the knowledges and skills would be applied. Often, the items presented in testing take the a form far removed from reality.

For example, in a test of word processing, a test-takers could be given a set of multiple choice questions that require knowledge of word processing to answer. It is clear that the test-taker is using word processing skills, but the tasks do not resemble how the act of word processing would takes place in real applications. A truly authentic measure would be to observe each test-take use a word processor to complete a real tasks several times over and rate their performance. While highly authentic, this is not realistic in large scale testing endeavors. A well designed CPEs can overcome some of the more egregious shortcomings of non-authenticity large scale testing by presenting tasks that elicit the same skills and behaviors that would be required to complete similar task in the reality in setting that constructed to resemble a natural setting. For example, in a

computer based test of word processing skills, a test-taker could be given a series of formatting task that must be completed ion a real word processor in the testing center. While still removed from word processing in the real world, such a task does require the active use of the same skill required in complete a real task, the setting is very similar to what the test-taker would encounter in reality, and the interaction with the word processor and the output of the task are closely related to how a test-taker would interact with the word processor in reality.

This allows for an inference from the scores on such items back to the domain of skills or construct to be measured that is more direct. In the word processor example given above, a work sample created on a word processor is submitted to assess word processing skills. The link between the task and the skills to be measured is clear and direct, and thus allows a strong inference between the scores on the CPE and the true ability of the test-taker. In short, the advantage, from a measurement perspective, of including CPEs in a large scale, high stakes examination is that they offer the potential for stronger validation arguments to be constructed by using high-fidelity items.

Challenges in Complex Performance Exercises

Using CPEs to build an assessment, or including some CPE items along with a selected-response (SR) section (Yao & Schwarz, 2006), certainly presents substantial challenges. Hardy (1995) links many of the practical challenges of CPE back to cost, in terms of financial expenditure and time. By their very nature, CPEs are more difficult to develop than SR items (Bennett, Jenkins, Persky, & Weiss, 2003). First, careful

consideration must be given to the domain of knowledge skills and abilities that a set of CPEs is to measure. After the skills of interest are defined, tasks must be developed that illicit the desired behavior in an authentic and realistic context (Miller & Linn, 2000; Wiggins, 1993).

Challenges in Task Development

Deciding what to measure and how to quantify the responses is a non-arbitrary task. Traditional performance assessment such as an essay or oral presentation must use detailed rubrics to measure information that directly relates to the skills and/or knowledge as defined in the construct domain. The introduction of complex performance assessments further complicates this process. For example, in a multi-step problem or in a simulation item, test-takers spend considerably more time reading or digesting the problem at hand, leading to a low data-collection-to-time-spent ratio (Bennett et al., 2003; DeVore, 2004). Test developers must be careful to allow for a significant number of scored tasks—or measurement opportunities (MO)—each of which relates directly back to the specified construct domain. If traditional PA and CPE are used in a large-scale testing format, the above concerns must be addressed with a process that allows mass production of these item types.

Technological Challenges

If the examination is delivered in a computer-based test (CBT) format, as many of these examinations are, technological issues may also affect development costs. CPEs

are more complex than SR items, so naturally these items carry with them many more components than the stem and distracters of SR items. Data structure that can store and manipulate the multiple pieces of a CPE must be in place to administer the test in a CBT format. Developing a framework for rendering the items on the computer screen is also of great importance. This framework must deliver items to test-takers in a manner that is authentic (items must look like tasks in the real world that would require these skills) and that encourages realistic test-taker/item interaction At the same time, the system underlying this process must capture the responses to each CPE task in meaningful way (Drasgow, Luecht, & Bennett, 2006; Luecht & Clauser, 2002).

Challenges in Scoring Complex Performance Assessments

Scoring expenses, particularly if items/tasks are scored by humans, must also be a consideration when CPE are used (Hardy, 1995; Wainer & Thissen, 1993). In large scale-testing, employing enough raters to score responses in a timely manner is a costly venture. In addition to the expense incurred at using human raters, there is also concern about the precision, stability and reliability of score that human raters produce (Clauser, Kane, & Swanson, 2002; Clauser, Swanson, & Clyman, 1999). Recent research has shown that explicit rubrics, extensive rater training, and a systematic reconciliation process to resolve discrepant ratings can help curb the inaccuracies of human scored items, though each of solutions further increases the cost of scoring and some degree of rating error is still inevitable (Clauser, 2000).

These considerations encourage the exploration of automated CPE scoring as an alternative to human response scoring. Automatic scoring has the advantage of reducing the cost of hiring and training human scorers while producing highly reliable and stable scores. Computerized scoring has also been applied successfully to essays (Burstein, Kukich, Wolff, Lu, & Chodorow, 2001; Y.-W. Lee, 2001; Rudner & Gagne, 2001; Sireci & Rizavi, 2000), user created graphical displays(Bridgeman et al., 1999), multi-step mathematics problems (Bennett et al., 2000; Bennett, Steffen, Singley, & Morley, 1997), and case-based simulations(Clauser et al., 2002; Clauser et al., 1999; Clyman et al., 1995). In all reported cases, automated scores compare favorably to human rendered scores, though substantive questions remain unanswered about the limitations of what particular aspects of a performance can be successfully scored (e.g. can the content or creativity of a response to an essay prompt be scored as well as the grammatical structure of the response?).

Two general families of automated scoring dominate the industry: rule-based scoring and algorithmic scoring. A rule-based system employs a set of Boolean rules to operationalize the actions that a rater might use in evaluating responses (Luecht, 2001, 2005). Algorithmic scoring, often the approach to scoring essay or spoken responses, requires that a sizeable sample test-taker response be human scored. This scored sample is processed by a computer, which systemically creates an scoring algorithm that recapture the human scores on any number of criteria (e.g. fluency, vocabulary, structure, etc.) as closely as possible. Once an algorithm has been developed, it can simply be applied to un-scored responses to produce scores. A rule-system could be applied to

render a partial credit score for a set of items, but is more often used to evaluate each separate MO and determine whether the task is accomplished or not, thus rendering a dichotomous score. This method has been employed in a number of certification/licensure settings that use simulation type items, and has been shown to produce scores that are comparable to human raters.

Statistical and Measurement Challenges

In addition to the cost and operational considerations, CPEs in mixed-format test also introduces several statistical complications. CPEs (which are often added to an existing assessment to explicitly measure something that SR items cannot) might measure a different trait than the SR section, and hence the resulting test can be considered multi-dimensional to some degree. Because the MOs of a single CPE are often associated with a common scenario or set of directions and resources, the items share a common contextual setting. Many CPE consist of sets of ordered or multi-step tasks. This can lead to chains of items where the scoring of correct or incorrect one task may directly influence the score of the related tasks that follow. Further, if a set of Boolean rules are used to automatically score the CPEs, these rule may allows explicit relations among the scored responses.

Because CPEs in mixed format test may constitute a separate dimension and CPE tasks may be linked by common contexts or scoring methods, any number of complicated statistical dependencies could be present. Significant item dependency and dimensionality can directly affect the estimation of test reliability, estimation of item and

ability parameters, estimation of test information, equating processes, and DIF detection. The validity of the scores produced by a test is also questionable if scaling and scoring methods do not adequately address the issue of dimensionality. As most tests that include CPEs are large-scale and high-stakes, it is essential that the scores produced are accurate and valid.

Purpose and Rationale of Research

This study will explore the statistical complications that are encountered when complex performance exercises (CPEs) are included with a sizeable selected-response (SR) section. Specifically, it explores the effects that several scoring and scaling procedures have on the residual covariance structures, (which are indicators of the statistical dependencies due to context/scoring and/or dimensionality), of such exams. The scoring aspect of this study compares analytically based dichotomous scoring strategies applied to CPEs with polytomous scoring strategies. The scaling aspects refer to two IRT scaling strategies: (1) scaling the entire test (SR items and CPEs jointly calibrated); (2) scaling the SR items and CPEs separately. This focus on different scoring and scaling strategies is directly relevant to many existing tests that employ some combination of SR items and CPEs

Delimitations of the Research

As this study represents a first exploratory step in understanding how the sources of residual covariance (evidence of statistical dependencies) manifest themselves in mixed

format exams that include complex performance assessments, a carefully defined context was defined for the investigation. Two general dimensional structures were considered: (a) a unidimensional set of items (i.e., both SR items and CPEs measuring the same common trait); and (2) an exam with two primary traits—a SR trait and a CPE trait. In the two-dimensional conditions, "simple structure" was assumed (Thurstone, 1947). That is, it was assumed that CPE items were only associated with the CPE trait, and SR items only associated with the SR trait. This study applies only to CPEs that are evaluated with a rule-based automated scoring routine that rendered dichotomous scores for a set of presented tasks.

CHAPTER II

REVIEW OF THE LITERATURE

The review of the literature is presented in three parts. The first section presents an overview of scaling methods and models that are frequently used in large-scale testing endeavors. Specifically, unidimensional item response theory (IRT) models, for both polytomously and dichotomously scored data, will be discussed with a detailed description of the models that will be employed in the current study. This section will also address various approaches to scaling when the exam is more complex—e.g. multidimensionality, items nested in testlets, etc.

The second part of the literature review will focuses on the assumption of local independence of items that accompany most measurement models. Local item independence and local dependence (LID) will first be defined and the relationship of LID to residual covariance explored. The sources of LID, in the context of the current research, and the consequences of ignoring LID due to these sources will be examined at length. Statistical methods for detecting LID, with a focus on Yen's Q_3(Yen, 1984), and studies from the literature that applied these methods are also discussed. Last, the second section will discuss the methods traditionally used to deal with LID. The last section of chapter two will summarize the previous two sections in the context of the current research and present the research question the study will address.

Scaling in Standardized Testing

In the context of psychological or educational testing, measurement can be defined as the systematic process by which numbers are assigned to individuals to represent properties or characteristics of those individuals (Allen & Yen, 1979; Lord & Novick, 1968). The last hundred years of testing have witnessed the development of a number of measurement models (e.g. classical test theory, generalizable theory, binomial error models, and most recently IRT) for just these purposes. The process of scaling is a framework within measurement where mathematical techniques are applied to determine what numbers should be used to represent the different amounts of a property that is the object of measurement (L. L. Thurstone, 1928; Torgerson, 1958).

In classical test theory (CTT), a scale is most often defined as the range of possible score points, and an individual placement on this scale is determined by summing each individual's responses to the items. In CTT, each item can be characterized in terms of difficulty by finding the proportion of all test-takers who correctly responded to the item (p-value), and discrimination by looking at how performance of the responses to an item relates to the overall scores (Lord & Novick, 1968).

CTT provides an intuitive and relatively simple method for scale creation, but is not without limitations. The characteristics of items are completely confounded with the measures of abilities. Take for example an item that is given to a very able population. The p-value would be very high, and one would conclude the item was very easy. The same item given to a less able population would have a lower p-value, thus giving a

different picture of the item difficulty. Likewise, a person's score on a test reflects the difficulty of the set of items that is administered. A difficult set of items would lead to a low score for a person, and give a low-estimate of ability. The same person given an easy set of items would lead to a higher score, and a higher estimate of ability. In CTT, the standard error of measurement (SEM) is the same for all examinees, regardless of their relative placement on the scale, and how that scale is informed by the difficulty of items. Finally, CTT has no method to predict how a test-taker of a given ability might respond to a given item. Last, CTT has failed to produce adequate solutions to practical testing problems (e.g. equating, detecting differential item functioning, test form assembly, etc.)

Item Response Theory

The development of item response theory (IRT) was largely motivated by a desire to address the weaknesses and limitations of CTT. IRT consists of a set of mathematical models that can be applied to raw data to produce a latent ability/proficiency parameter for each test-taker (in itself a scale), as well as a set of parameters that describe the properties of each item (Hambleton & Swaminathan, 1985; Lord, 1980). These models began to take form in the late 40's and earlier 50's under measurement theorists like Tucker(1946), Lord(1952), Rasch (1960), and Birnbaum(1968).

The earliest IRT models dealt exclusively with unidimensional tests and dichotomous responses. Models to produce scales when the data is polytomous (Masters, 1982; Muraki, 1992; Samejima, 1969) were developed in the following decades. Most

recently, IRT models that can model tests that measure more than one ability (multidimensional) have also been developed for use (McDonald, 1981, 1982b; Reckase, 1985). Initially, even the simplest of these models was too computationally complex for practical application, but the advancement in computer technology in the 70s and 80s provided the opportunity to use IRT models in applied test settings. In the last three decades, IRT has become the field standard for scaling large-scale testing data.

The collection of IRT models is by definition a collection of scaling methods. Using the response data, characteristics of the items and the latent ability of individuals are estimated simultaneously and with respect to one another. The latent score(s) constitute(s) the scale(s). Often, the convenient assumption is made that the latent variable(s) are distributed as univariate or multivariate normal. The performance of test-takers on a given item can be explained by a mathematical function (the item response function or IRF) that states that as an individual's latent ability increases, the probability of a correct response also increases (i.e., the function is strictly monotonic). If an IRT model fits the empirical data, then in theory, the two sets of parameters will be locally independent and invariant across population subgroups and measurement conditions (Lord, 1980). Rasch (1960) referred to this invariance principle as "specific objectivity." Invariance implies that if the same set of items are given to two separate groups, the same item parameters, ignoring random errors of estimation, would be obtained. Likewise, two individuals with same ability from two separate test administrations will have the same probability of a correct response on the same item. This invariance

property and is one of the cornerstones of IRT and a major distinction from CTT (Hambleton & Swaminathan, 1985).

Even the most basic of IRT models assumes that the set of items collectively measure a single trait. That is, other than random measurement errors, an individual's ability or proficiency—a single latent trait—will explain all the variance in an individual's responses. This is a fundamental assumption when employing any unidimensional IRT model (Lord, 1980). This assumption is also related to the concept of local independence of items. Local independence means that, after taking into account an examinee's ability (or abilities in the multidimensional case), no residual association exists between the examinee's responses to different items. In other words, local independence will hold if, and only if, the modeled ability (or abilities) is (are) the only factors that contribute to (explain) a test-taker's responses on all items. In the unidimensional case, a single ability or proficiency, denoted θ, completely represents the latent space. In the multidimensional case, the latent space includes two or more abilities, $\boldsymbol{\theta}$. If the assumption of unidimensionality holds for a unidimensional IRT model applied to real data, the assumption of local independence follows logically (Lord, 1980; Lord & Novick, 1968).

IRT Models for Dichotomous Data

The earliest IRT model to gain notoriety was the two parameter normal ogive model as presented by Lord (1952; Lord & Novick, 1968). The logistic family of IRT models, as parameterized by Birnbaum (1968) and Rasch (1960), are more

mathematically tractable than their ogive counterparts, and so usurped the ogive models in the operational practice. The IRF of the three parameter logistic model (3pl), one of the most frequently used IRT models, is defined by:

$$P\left(X_j = 1 \middle| \theta_i, a_j, b_j, c_j\right) = c_j + \left(1 - c_j\right)\frac{1}{1 + e^{-Da_j\left(\theta_i - b_j\right)}} \tag{2.1}$$

where θ_i represents the latent ability of the i^{th} examinee, $P(X_j=1|\theta_i, a_j, b_j, c_j,)$ is the probability that the i^{th} examinee will correctly responds to the j^{th} item. The parameter, b_j, is the difficulty or threshold parameter and describes the value of θ where a test-taker has equal chance of correct and incorrect responses. The discrimination parameter for the j^{th} item, a_j, is proportional to the slope of the IRF at the threshold and is the scaling weight of the item. The parameter c_j, is often referred to as the "guessing" or "pseudo guessing" parameter and represents the lower asymptote of the IRF. The value D, is a constant scaling factor that places the scale of the latent ability on the standard normal metric when set to 1.702 or on the logit metric when set to one. Constraining the c parameter to be zero in equation 2.1 reduces the 3pl model to the commonly used two parameter logistic model (2pl). Further constraining all of the *a* parameters to be equal (or in the case of the Rasch model set all *a* parameters to a value of one) reduces the model to the one parameter logistic model (1pl). The expected response on any item j, for any test-taker *i*, can be modeled by

$$E\left(x_{ij} = 1\right) = P\left(X_j = 1 \middle| \theta_i, a_j, b_j, c_j\right) \tag{2.2}$$

where $P\left(X_j = 1 \middle| \theta_i, a_j, b_j, c_j\right)$ is the probability of a correct response as defined in equation 2.1.

IRT Models for Polytomous Responses

The need for IRT models for items with multiple score categories stemmed largely from the fact that these items exist and are frequently used in psychological and educational assessment. Polytomous items also have substantive appeal as well. In a general sense, the addition of more than two score categories allows for measurement information across a wider range of the ability scale (Ostini & Nering, 2006). Samejima (1976) demonstrated the increase in statistical information when using a polytomous rather than a dichotomous model. Masters (1988) also has described the increase of diagnostic information made available about test-takers by adding score categories.

In the context of CPEs, the advantage of polytomous IRT models is clear. Performances like essays and interviews, if they are to be rated by humans, often rely on complex rubrics that result in a rating that can fall into a range of score categories. Likewise, automatic scoring algorithms or rules applied to these items could be developed to produce scores in any number of score categories. Items that require multiple steps could be scored in such a manner that partial credit could be awarded. Additionally, as will be addressed in the next section of Chapter Two, polytomous models have also been employed as a potential method for dealing with sets of related dichotomously scored CPE items.

Two families of ordered-category polytomous models are frequently used in practice (Thissen and Steinburg, 1986). The first family of polytomous IRT models were developed are based largely on a framework defined by Samejima (1969) under what she termed the graded response model (GRM). Thissen and Steinburg refer to the family of GRM models as "difference models", since the successive response probability functions model the likelihood of a cumulative observed score as a function of the underlying trait (e.g., $Pr\{x \geq 2|\theta\}$). To derive the probability of a particular score (e.g., $Pr\{x=2|\theta\}$, we subtract the cumulative probability of obtaining the next higher score from the probability of obtaining the current score (e.g., $Pr\{x=2|\theta\}=Pr\{x \geq 2|\theta\}-Pr\{x \geq 3|\theta\}$). The second family of models, termed "divide-by-total" models by Thissen and Steinburg, were founded on the multinomial probability function, where each score category probability function is essentially normalized by the sum of the category probability functions over the entire response space. The earliest of these models were based on the Rasch exponential family of models (Andrich, 1978; Masters, 1982).

The expressed intent of both families of models is to describe the probability of a test-taker falling into a given score category given her/his ability. To accomplish this, both families of models employ a set of dichotomizations of the response categories. The dichotomizations take place at the boundaries between two adjacent categories and the model attempts to find the probability that a test-taker will be placed in the higher category. These distinctions are modeled as a dichotomous IRT model and the resulting probabilistic functions describing the dichotomies at each boundary are called category boundary response functions (CBRF). Ultimately by combining all the CBRFs in a

principled manner, the probability of falling into a specific category can be obtained. It is the method and the form of this dichotomization and the resulting CBRFs that distinguishes the families of models.

Samejima's graded response model dichotomizes the responses globally. This means the dichotomy of interest when describing the boundary between categories two and three in a five category item would be characterized by considering all the responses three or greater versus the responses falling into a category below 3. The resulting CBRF would then describe the probability of scoring in category three of greater, P_3^*. Finding the probability of falling exactly into a given score category can be simply defined as finding the difference between the upper and lower CBRF for that category—or $P_k = P_k^* - P_{k+1}^*$, where k is the score category of interest. This type of model is referred to as a difference model (Thissen & Steinberg, 1986).

Divide-by-total polytomous models (i.e., Rasch based models) use a localized approach to dichotomizing the response categories. The modeling of each CBRF is based only on splitting the responses only along the two categories immediately adjacent to the boundary of interest. This process allows the response functions to cross, within any CPE. For example, the functions $\Pr\{x=2|\theta\}$ and $\Pr\{x=3|\theta\}$ may cross, thereby sharing the same probability space with respect to θ. The normalization inherent in the "divide-by-total" modeling of the response function produces a proper probability response function.

While both families of models are useful in applied applications, Ostini and Nering (2006) suggest that the divide-by-total approach to modeling category boundaries

allows for a decomposition of discrimination into both an item and category component, where Samejima models allow only the item discrimination to vary. This flexibility may allow for more complex representation of the polytomous item.

This study employs Murkai's (1992) Generalized Partial Credit Model (GPCM), to scale all polytomous data, which is most general of the divide-by-total models for polytomous data. This model allows for the flexibility attributed to the Rasch models, but unlike the other Rasch models, also allows for a separate slope parameter to be estimated for each item. The item category response function (ICRF) is defined as:

$$P_{ik}\left(X_j = k \middle| \theta_i\right) = \frac{e^{\sum_{v=1}^{k} z_{iv}(\theta_i)}}{\sum_{c=1}^{K} e^{\sum_{v=1}^{c} z_{iv}(\theta_i)}} \tag{2.3}$$

Where

$$z_{jk} = a_j\left(\theta - b_{jk}\right) = a_j\left(\theta - b_j + d_k\right) \tag{2.4}$$

In equation 2.3, a_j is a slope parameter representing the discrimination of each item j. The parameter b_{jk} is an item-category parameter for the k^{th} category of the j^{th} item. The parameter can be decomposed into b_j, an overall location parameter for item j, and d_k , a category location parameter (with respect to the overall location of the item) to for the k^{th} category of the j^{th} item. $P_{ik}\left(X_j = k \middle| \theta_i\right)$ is the probability that the i^{th} person of ability θ_i will score in the k^{th} response category on the j^{th} item. Equation 2.3 takes the form of a divide-by-total model, where the numerator describes the likelihood of somebody at the

given trait level is placed in the higher category at each boundary dichotomization up to the category of interest, k (Thissen & Steinberg, 1986; Thissen, Steinberg, & Mooney, 1989). The denominator is the sum of all numerator values for ALL possible score categories.

The expected response on any item j, for any test-taker i, can be defined using:

$$E\left(x_{ij}=k\right)=\sum_{k=1}^{K} y_{jk} P\left(x_{ij}=k \mid \theta_i, a_j, b_{jk}\right) \tag{2.5}$$

Where K is the total number of score categories, y_{ik} is the weight (typically the same as the category values) for the k^{th} category, and $P\left(x_{ij}=k \mid \theta_i, a_j, b_{jk}\right)$ is the probability that a person with ability θ_i falls into response category *k*. Muraki (1992) demonstrated that the GPCM can be conformed to function as other Polytomous Rasch IRT models (i.e., the partial credit model, the rating scale model) by constraining the item parameters.

Extensions of Unidimensional IRT Methods

In practice, response data rarely conforms exactly to the rigorous assumptions of unidimensional IRT models. A failure to fully define the latent space will lead to violations of local independence, and in the case of unidimensional IRT, violations of the assumption of unidimensionality. If a single dominant latent ability is presumed to explain most of the variance in the response data, a weaker set of assumptions, essential independence and essential unidimensionality, may be tested to verify that a unidimensional IRT scaling model is appropriate (Nandakumar, 1991, 1993; Nandakumar & Junker, 1993; Stout, 1990). If such a claim cannot be verified or is not

deemed appropriate, other IRT models that allow for more than one factor or dimension could be employed.

Up to this point the assumptions of local independence and unidimensionality have been discussed synonymously. Both assumptions speak to appropriately accounting for the complete latent space. Two types of factors may contribute to the latent space, factors or traits that are relevant to the construct or domain that a test is intended to measure, and factors are irrelevant to the construct (nuisance factors).

The main distinction between these types of factors is one of valid score reporting. "Nuisance factors" are typically induced through some artifact of the test administration process (item topics, method of assessment, etc.)[1]. Nuisance factors do not constitute any trait of interest, regardless of whether the nuisance factor is related to the primary construct. Therefore, we do not want the nuisance factor(s) to affect the scaling procedures that ultimately produce the ability or proficiency estimates for each examinees. Nonetheless, if these factors influence a test-taker's response and an IRT model does not account for them, several of the key assumptions of IRT may be violated: the assumption of local independence and the assumption of unidimensionality. Depending on the magnitude (influence) of the nuisance factor(s), there could be dire impact on the estimation of item parameters and latent ability scores. Furthermore, the validity of the score produced in general must be questioned if relevant traits are being "contaminated" in the scaling process; that is, if some combination of the intended traits

[1] It should be noted here that this prospective on what constitutes a "nuisance factor" represents one, though perhaps the dominant, perspective in contemporary psychometrics. Others, like Kane's (Kane, 2006; Kane, Crooks, & Cohen, 1999) trait theory, propose that many methods factors, context of testing, and other traits (see figure 2.2 in Kane,2006) that have been traditionally labeled nuisance factors actually constitute a viable part of the trait, and directly influence the target domain of skills that are to measured.

and the nuisance factor(s) may be manifest in the responses of certain examinees, or the population at large.

In short, the latent space must be reasonably identified and accounted for in order to produce valid score estimates. This means that appropriate models must be applied to the response data to extract and generate a scale for the underlying traits of interest, as well as factoring out any residual variance that might be attributable to nuisance factors. Two multidimensional IRT approaches to dealing with multiple intentional traits are discussed in the next section. Models designed to account for the nuisance factors (and their influence on the response data) follow.

Multidimensional Item Response Theory

The use of a multidimensional IRT (MIRT) model is only useful and appropriate if the item responses are dependent on more than one latent factor, AND more than one factor is deemed relevant to the construct. McDonald (1981, 1982a, 1982b, 2000) first characterized MIRT as a nonlinear factor analysis of items. (also see Bock and Liberman, 1970). More typically, these models are presented as general forms of the familiar unidimensional logistic IRT models (Reckase, 1985, 1997; Reckase & McKinley, 1991). Several MIRT models have been developed (van der Linden & Hambleton, 1997), but as many of these model can be seen as extensions or generalizations the compensatory MIRT model for dichotomous data and no MIRT models are applied in this study, attention here will be restricted to the one model for descriptive purposes only. For p dimensions, the IRF is defined as:

$$P\left(X_j = 1 \middle| \boldsymbol{\theta}_i, \mathbf{a}_j, d_j, c_j\right) = c_j + (1 - c_j)\frac{e^{\left(\mathbf{a_j\theta_i} + d_j\right)}}{1 + e^{\left(\mathbf{a_j\theta_i} + d_j\right)}} \quad (2.6)$$

where θ_i is a p x 1 vector of latent abilities for the i^{th} person, a_j is 1 x p vector of discrimination parameters that describe the level of association of item j to each of the p dimensions, d_j is a threshold parameter for item j, and c_j is a lower asymptote parameter. The obvious advantage of such a model is that ability estimates for all p of the traits of interest can be estimated. However, MIRT models are not without practical and statistical limitations. Estimation of these models requires large samples that may not be practical in many testing situations. More importantly, using a complex MIRT model introduces complications involving factor rotational indeterminacies and parameter identification issues (Luecht, Gierl, & Huff, 2006). Such a scaling would at best produce item and person parameter estimates that are only locally useful (i.e. interpretable only for scaling of a single form or administration) or worse, lead to inconsistent score interpretations and test equating over time. Luecht, Gierl, and Huff offer a strong cautionary note that if the nature of dimensionality is not understood and controlled in test development, then it is not desirable, likely not understood, and hence, should not be used to produce multiple ability estimates for reporting.

Principled Multidimensional Information

As it is likely that any construct worth testing is multi-faceted and too complex to encapsulate in a single latent trait , Luecht et al (Luecht et al., 2006) have encouraged the

treatment of dimensionality to be largely addressed proactively in test development, rather than in a post-hoc factor analysis of the data. Luecht et al. (2006) call for the development of principled multidimensional information (PMI)—through empirically informed, carefully engineered assessment tasks, not just allowing idiosyncratic multidimensionality to emerge in the test data. This approach to test development—which is called "assessment engineering" (Luecht, 2006, 2007a, 2007b), allows some of the more tedious statistical complexities of the MIRT models to be resolved through careful, empirically verified task and test design principles, rather than via tautological statistical assumptions and arbitrary manipulations of model-based constraints that arise when a complex models are applied to extant data.

The AE approach to test development is meant to produce multidimensional tests that exhibit what Thurstone (1947) termed "simple structure" (as a criterion for factor weight matrix rotation). McDonald (1999) offered a comment of similar intent regarding the common factor model. McDonald states that if a few stringent criteria are met, this model is guaranteed to be identified. First, for any factor there are at least three measures with non-zero factor loadings that have zero loading on all other factors. If a factor has only two measures with non-zero factors loadings but with zero loadings on all other factors, the model can be identified if this factor is correlated with other factors in the model. At its most conservative interpretation, McDonald is calling for simple structure, with or without correlated factors, in a common factor model to assure that the model is identified.

Since any MIRT model can be viewed as a particular parameterization of a common factor model (see McDonald, 1999), it is clear why a post-hoc MIRT scaling may have trouble meeting McDonald's criteria for identification. If dimensionality is considered only after the items are developed and the test is administered, the number of dimensions must determined empirically (Stout, Habing, Douglas, & Kim, 1996) or by experts, and a pattern matrix describing item-trait relationships could be retro-fitted based on expert opinion or based on an exploratory analysis. In an exploratory analysis, the various choices of rotational criteria applied to the data matrix can change both the pattern and magnitude of discrimination parameters and the correlations between traits. If a confirmatory approach is taken to meet McDonald's criteria, but the true structure is not known (even approximately) *a priori*, and simple structure is artificially imposed, the correlation between traits will be poorly estimated and the model fit may be very poor.

A well-implemented AE approach to designing tests that exhibit PMI for a particular population of examinees specifies the number of traits of interest and builds independent families of items to measure only a single specified trait—that is, simple structure is established by design (Luecht, 2006, 2007a, 2007b). By using carefully designed task models item templates, items measuring the same traits in the same way can be produced in mass. This means the same underlying dimensional structure can be created over various forms over time. The assumption is that the relationship of these operationalized traits would also be stable over time. The clear advantage is that by establishing simple structure and creating the same relationship between traits over time allows the rotational indeterminacy to be resolved if the assumption of local

independence holds within each independent cluster (i.e. the single ability trait for each cluster is the only factor needed to explain response).

In practice, a test developed as described above would consist of multiple unidimensional scales and can be scaled using unidimensional IRT models (if local independence holds within each scale) rather than more complex MIRT models. Any relationship between traits can be described by the correlations between the multiple scales. An AE approach to test development must be coupled with extensive confirmatory studies that evaluate the item templates, item design, test construction, and ultimately the assumptions of the scaling methods (i.e. local independence) employed.

Other Complex Models

Often the factors that cause violations of local independence are not traits that relate to the construct of interest. In these situations there is no reason to apply a multidimensional IRT scaling approach, as the extraneous dimensions constitute systemic noise rather than measurement information. These are termed "method factors" in some of the measurement literature (Campbell & Fiske, 1959). Methodological approaches have been developed for these situations, and two of the major modeling approaches related to IRT scaling are presented below.

One family of these models is designed to deal with context-dependent sets of items. Take, for instance, a reading test that consists of reading passages with clusters of related items. The primary trait of interest is the reading ability of each test taker, but a test-taker's familiarity with the topic of one or more reading passages may contribute to

the probability of correctly answering associated questions. These related sets of items constitute a "testlet", and introduces a testlet factor(s) into the defined latent space. The effects of each testlet factor on the responses of the test-takers is a nuisance factor that needs to be accounted for in order to report scale scores for the primary trait, reading.

Bradlow, Wainer, and Wang's (1999) random effects testlet model and the generalizations of this model by Li, Bolt, and Fu (2006) have been shown effective in partialing out the effects of testlets. Demars (2006) also illustrates how a bi-factor model (Gibbons et al., 2007; Gibbons & Hedeker, 1992) can be successfully applied to assessments that contain testlets. A full discussion of models appropriate for testlets is beyond this study, and interested readers are referred to the extensive studies present in the literature for further information.

Local Item Dependency

Local independence of item responses is a notion that appears in numerous forms in test theory. In classical test theory, it is assumed that errors of measurement are uncorrelated, given the true score of an examinee (Lord & Novick, 1968; McDonald, 1999; Yen, 1984). In item response theory (IRT), a set of items is considered locally independent with respect to the assumed model if, after conditioning on an examinee's proficiency, the joint probability distribution of all items is equal to the product of the univariate probability distributions of each item (Hambleton & Swaminathan, 1985; Lord, 1980). Formally, this is the strong definition of local independence and is stated mathematically in equation 2.7

$$P\left(\mathbf{X}_i \middle| \theta_i\right) = \prod_{j=1}^{J} P\left(X_{ij} \middle| \theta_i\right) \tag{2.7}$$

Where X_i is the vector of observed responses to J items for test-taker i, and X_{ij} is the response of test-taker i to the j^{th} item. In practice, a weak definition of local independence is often used to investigate the appropriateness of this assumption. Weak independence states each item pair's joint distribution is equal to the product of corresponding marginal distributions after accounting for each subject's ability.

$$P\left(X_i, X_j \middle| \theta\right) = P\left(X_i \middle| \theta\right) P\left(X_j \middle| \theta\right) \text{ for } i \neq j \tag{2.8}$$

Weak local independence is a less stringent requirement that is necessary but not sufficient for strong local independence (Stout et al., 1996). However, it is reasonable to assume that if variables are pair-wise independent, higher order dependencies, though possible, are highly implausible (McDonald, 1997). If equation 2.8 holds for all item pairs, the trait proficiency (θ) accounts for all of the information relevant for each examinee, thus allowing the items to be evaluated independently (Yen, 1984).

This idea is easily expressed in terms of conditional covariance as well. If items X_i and X_j are locally independent they will have a covariance of zero, after conditioning on some ability θ.

$$\text{cov}\left(\mathrm{X_i}, \mathrm{X_j} \middle| \theta\right) = 0, \text{ for } i \neq j \tag{2.9}$$

A definition of local item dependence (LID) follows logically from the definition above—the presence of conditional covariance between a set of items.

$$\mathrm{cov}\left(\mathrm{X_i},\mathrm{X_j}\middle|\theta\right) \neq 0, \text{ for } i \neq j \tag{2.10}$$

That is, conditional covariance in either a positive or negative direction indicates that performance on one item is related to the expected performance of the examinees on the other item (Rosenbaum, 1984). This can also be expressed in factor analytic terms: LID is present if, after extracting the first factor (roughly equivalent to conditioning on the trait of primary interest θ), there is a non-zero residual covariance between some items (McDonald, 1982b). These non-zero covariances indicate that there may be one or more additional factors that explain the remaining variance (Yen, 1993). The additional factors are the potential sources of LID and may or may not be vital to the trait or behavior that is being measured.

Sources of Local Item Dependencies

Yen (1993) lists several potential sources of LID in testing situations. These sources include (1) external assistance or interference in the test taking process, (2) test speededness, (3) the effects of fatigue in lengthy test settings, (4) practice affects, (5) passage or contextual associations, (6) associations due to scoring methods or rules, and lastly, (7) subsets of items that require different knowledge, skills and abilities. This section will outline the potential causes of LID that are of particular importance in performance assessment—specifically, dependencies due to context, scoring, and dimensionality.

A context-dependent set of items typically consists of a common stimulus shared among several items that require the use of the stimulus in the response process

(Haladyna, 1992). A classic example of context-dependent items is a set of reading comprehension items linked to a common text passage. In this setting, test-takers' familiarity with the topic as well as overlaps in the specific information used to respond to multiple questions may differentially impact the pattern of correct responses across test-takers. Sireci et al. (1991) used several reading passages from the SAT verbal section to demonstrate the degree to which LID existed and affected the reliability estimates of these tests. Lee (2004) demonstrated similar amounts of LID between items within a passage dependent set in an English as a foreign language (EFL) examination.

Context-dependent item sets are not limited to dependency on a reading passage or in the context of a reading comprehension examination. Ferrara, Huynh, and Baghi (1997) demonstrated that substantial amounts of LID are observed in sets of math problems that are linked to a common theme or stimuli. Yan (1997) and Ferrara, Huynh, and Michaels (1999) demonstrated that science assessments linked to a common experiment, graphic, table, or general topic tend to display LID that is due in part to contextual considerations. Similar patterns of LID attributable to context related passages in medical examinations (Zenisky, Hambleton, & Sireci, 2002) and certification examinations for accountants (Goodman & Luecht, 2007) have also been documented. As many complex performance assessment share a common setting, stimuli, set of directions, or set of resources, some LID due to shared context can be expected in these items (Wang, Cheng, & Wilson, 2005).

Item-level scoring procedures, particularly on performance items, can lead to LID between items. For example, a set of rules for automated scoring may share objects

among rules (i.e. awarding credit in more than one place for a correct response on a particular item) and lead directly to associations between items. Items chains, a series of items where a test-takers' responses depend directly on the responses given to previous task (e.g., a multi-step CPE task) or requiring a test-taker to explain a previous response, can create associations between items. Both scoring and contextual dependencies are considered to be nuisance factors that do not have direct implications on measured traits or skills, but when left unaccounted for, can have adverse effects on the properties of the scores produced.

If more than one skill or trait is required to successfully explain an examinee's response, a test can be considered to have some degree of multidimensionality. LID is an indicator that multiple proficiency traits may underlie the collective response patterns for a set of items on a test. Ultimately, additional dimensionally relevant score scales may be needed, if the practitioner decides that the skills they represent are essential to the measurement purposes of the test.

Consequences of Ignoring Local Item Dependencies

Ignoring LID, regardless of its cause, affects the psychometric properties of tests. For example, if a unidimensional IRT model is fit to the data, but then response patterns are encountered that violate the assumption that items are conditionally independent, test information and reliability are overestimated and the standard errors of the ability estimates are underestimated (Chen & Thissen, 1997; Sireci & et al., 1991; Yen, 1993). Reese (1995) describes low-score underestimation and the high-score overestimation in

sets of items that exhibit LID. Further, LID is known to affect the estimation and accuracy of item parameters. Wainer and Wang (2000) found that lower asymptotes were overestimated when dependencies were ignored between testlets. Ackerman (1987) found that item discriminations were overestimated when set of items were locally dependent. When items are to be banked for use in automated test assembly or computer adaptive testing, inaccurate item parameter estimates can call the fairness of the test into question (Thompson & Pommerich, 1996). If residual covariances differ for various population subgroups, differential item functioning (DIF) results may be affected. Finally, test scaling and equating practices—which rely on accurate parameter estimates—can be adversely affected by LID (De Champlain, 1996; Reese & Pashley, 1999)

Assessment of Local Item Dependencies

Several methods for assessing LID have been developed. While most of these methods have been applied almost exclusively to dichotomously scored data, extensions for polytomously scored data have also been developed for many of these procedures. This section will explore some of the most successful methods for assessing LID.

The first two methods described, the Pearson χ^2 and the likelihood ratio test (G^2), use the observed and expected frequencies of score patterns for pairs of items to assess LID. Chen and Thissen (1997) applied both statistics to dichotomously scored data that was scaled with unidimensional IRT models. Expected response frequencies for each item pair were then predicted using the IRT model. Lin , Kim and Cohen (2006)

generalized both methods to apply to polytomously scored data. Both these statistics are distributed as χ^2 with $(K-1)^2$ degrees of freedom (where K is the number of score categories). Pearson's χ^2 is computed as:

$$\chi^2 = \sum_{i=1}^{K}\sum_{i=1}^{K}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{2.11}$$

where K is the maximum number of score categories, and O_{ij} and E_{ij} are the observed and model-derived expected values for the cells in the K x K table. The likelihood ratio G^2 is:

$$G^2 = \sum_{i=1}^{K}\sum_{i=1}^{K} O_{ij} \ln\left(\frac{E_{ij}}{O_{ij}}\right) \tag{2.12}$$

where the elements of this equation are defined in the same manner as in the Pearson χ^2 statistic. Both methods are effective in detecting dependent item pairs, but are limited to detecting the presence and not the direction of LID. G^2 has been shown to be slightly more powerful than χ^2 in detecting LID.

Conditional inter-item correlations have also been used to assess LID (Ferrara & et al., 1997; Ferrara et al., 1999). In this method, test-takers are sorted into a limited number of groups based on total test score, and inter-item correlations computed for each score interval. Items expected to display LID due to contextual, scoring, or underlying dimensionality are aggregated across the score levels to give a measure of LID. Items not expected to display LID are aggregated separately, and compared to this first measure.

The presence of LID can be tested by comparing separate reliability estimates of the same tests (Sireci & et al., 1991; Wainer & Thissen, 1996; Zenisky et al., 2002). The first reliability estimate assumes that all items are locally independent. The second estimate models the reliability after forming testlets for context-dependent sets of items. If the testlet-based reliability is substantially lower than the item-based estimate, LID is present for some or all of the items in the testlets.

If the source of the LID is presumed to be due to dimensionality, several methods are effective to either explore or confirm the dimensionality structure of a set of items. Zhang and Stout (Stout, Habing, Douglas, Kim et al., 1996; Zhang & Stout, 1999) use conditional covariance structures to assess the number of dimensions present and provide a statistical measure of fit in the DIMTEST and POLY-DIMTEST procedures. Nonlinear factor analysis provides a convenient and conceptual mechanism to characterize various knowledge or skill dimensions. Structural equation modeling (SEM) is used in a similar fashion to confirm a proposed dimensional structure, or to compare competing dimensional structures. SEM also provides a battery of fit statistics and residual matrices for assessing the degree to which the data fits a proposed multidimensional model (De Champlain & Gessarolli, 1995, 1997, 1998).

The Q_3 statistic (Yen, 1984) is the correlation between item pairs after accounting for some measure of performance, θ^*. In this sense Q_3 is a standardized residual covariance structure for all item pairs. If the assumption of local independence holds, and θ^* adequately represents the latent space, the item pair-correlations should be zero or, after accounting for θ^*, any residuals constitute random measurement error. One distinct

advantage of Q_3 is that it takes the form of a correlation. This simplifies the interpretation of the magnitude of LID present and also allows the direction of the residual covariance to be assessed. Formally Q_3 is defined by:

$$Q_3 = r\left(d_j., d_i.\right) \tag{2.13}$$

where

$$d_{ja} = x_{ja} - E\left(x_{ja}\right), \tag{2.14}$$

x_{ja} is the observed score on the i^{th} for the a^{th} examinee and $E\left(X_{ai}\right)$ is the expected score of the k^{th} examinee on the i^{th} item. Q_3 was initially developed for dichotomous items scaled with a unidimensional IRT model. In this case, $E\left(x_{ia}\right) = P_i\left(\hat{\theta}_a\right)$, where $\hat{\theta}_a$ is an IRT based estimate of latent ability of test-taker a. This statistic is easily adjusted for use when a polytomous IRT model is used, by simply redefining the expected score function to be $E\left(x_{aj}\right) = \sum_{k=1}^{K} y_{jk} P_{jk}\left(\hat{\theta}_a\right)$, as defined in equation 2.5.

Q_3 is an effective way to describe the presence and magnitude of LID and has been demonstrated to outperform other LID indices (Chen & Thissen, 1997; Habing, Finch, & Roberts, 2005; Yen, 1984, 1997). Most recently, it has been suggested that Q_3 can be generalized to address models outside of unidimensional IRT (Goodman & Luecht, 2007; Goodman, Luecht, & Zhang, 2008). These studies propose that the definition of the conditioning variable θ^*, be expanded beyond a single IRT latent trait

estimate to represent any combination of variables that best represent the latent space. In this context, θ^* could be a composite trait or a vector of traits produced from several separate calibrations, a MIRT model, or an alternative model (e.g. a bi-factor model, a model for testlets, etc.)

Because Q_3 includes an item score explicitly in x_{ja} and implicitly in $E(x_{ja})$ through the use of $\hat{\theta}_a$, values will be slightly negative due to part-whole contamination. The expected value of Q_3 when local independence holds for all item pairs is:

$$E(Q_3) = -\frac{1}{j-1} \tag{2.15}$$

where j is the total number of items used to estimate the latent score θ^* .

Strategies for Addressing Local Item Dependencies

In situations where LID is present, or likely to be present, due to contextual and/or scoring concerns, certain courses of action are advisable to reduce the effects and magnitude of LID. The most common solution is to form "testlets" from the related items and create one or more "super" polytomous items from the cluster by summing the individual scored objects. The resulting testlet-based super item can then be scaled using an IRT model for polytomous data. Polytomous scoring of testlets has been demonstrated as effective in reducing LID (Sireci & et al., 1991; Stark, Chernyshenko, & Drasgow, 2002; Yen, 1993; Zenisky, Hambleton, & Sireci, 2001). However creating polytomous items from unrelated subsets of items has been shown to decrease reliability and test information (Yen, 1993). If a test has several related sets of items (i.e. several

reading passages with related clusters), then this method is most effective if the created polytomous items can be created so that local independence is maintained across all the newly created polytomous items.

The course of action is less clear when a test exhibits multidimensionality. The simplest and perhaps most common practice is to continue to assume that the mixture of multiple dimensions forms an essentially unidimensional measure. In that case, the resulting total-test ability estimate can be shown to represent a weighted composite of some unknown number of traits. If the extent of multidimensionality is small and unrelated to specific features of the items or content of the test, this solution may be reasonable. The composite ability estimate is effectively weighted according to the relative number of items linked to each trait and the average information exhibited by those items. However, as the magnitude of multidimensionality increases, the projection of any ancillary dimensions onto a single reference composite can alter the nature of the total-test composite in unpredictable ways.

Another approach to dealing with multidimensionality can also be employed: scaling the sets of items assumed to represent different traits or skill sets separately (a PMI approach would further assume that the items were distinctly developed to measure predefined traits). This approach allows separate scores for each scale/dimension to be reported and, ideally, results in ability estimates that adequately explain the responses to related sets of items. Again, this course of action is not without practical consequences. Breaking the complete set of test items into separate tests will results in smaller tests. Smaller tests, in turn, yield less reliable scores, and less reliable scores may adversely

affect the quality of the ability estimates, themselves. Statistical augmentation can be used to improve the reliability of the multidimensional estimates, but not without regression bias to the mean. Finally, creating an appropriate and stable total-test, composite score (if one is needed) can become a tedious procedure from an equating perspective.

More complex scaling methods are also available when a test is likely multidimensional. A great number of multidimensional IRT (MIRT) models can estimate multiple abilities jointly, describe the relationship between sets of traits, and allow for factorial complex structures within the test. These scaling methods are more computationally complex and require much larger sample sizes. Furthermore, software packages to fit these models to data tend to be limited. Technical statistical issues such as rotational indeterminacy also remain largely unresolved for MIRT models.

Summary in the Context of Current Research

Dealing with LID is complicated by an intractable confounding of the interactions between subsets of examinees from the population and subtle or not-so-subtle characteristics of the assessment tasks. Simple statistical methods can help test for the presence of LID. However, conclusively determining the source or sources of LID is not straightforward. Take, for example, complex performance assessments. These items are typically developed to measure analytical, problem-solving, evaluative and communication skills that are difficult to measure in traditional SR questions. Whether these skills align with content to form an ordered continuum is open to empirical

investigation. Members of the populations may also differ in their opportunities to learn these new skills. Once items are written for a particular performance assessment, there are additional interactions that may involve prior knowledge for some groups with common contexts or stimuli used to link the items or MOs together. In addition, the items may be scored with complex scoring rules having various implicit or explicit associations among the components used in scoring. How those components are learned may differ for various population subgroups. In short, our capability to detect LID probably surpasses our ability to understand it.

The current study examines the effects that different methods for scoring CPEs and scaling tests that include CPEs have on the residual covariance structures of these tests. The scoring and scaling methods selected represent some of the approaches that might be employed when dealing with a test of this nature. The methods were examined over several test lengths, proportions of complex performance assessment items, and sample sizes that might be observed in practice. The residual covariance structures produced in the simulation study are partitioned into three sets of interest: the relationships of MOs within a single CPE, relationships between MOs on different CPEs with a single simulated test, and the relationships between the collective set of CPE MOs and the SR items within a single test. The residual covariances of item pairs are assessed in terms of their magnitude and their direction. In any of the three comparisons above, significant negative (divergent) covariance provides evidence that the items may be measuring different traits. Positive (convergent) covariance between two items may indicate that contextual or scoring associations are leading to dependencies (this is only

possible within a single CPE as only those items share scoring rules and contextual settings). Positive covariance is also evidence that the item pair may be associated with some other trait not represented in the scaling model selected.

Clustering related items into sets and forming polytomous items from these sets is one potential way to control the amount of LID due to a common context or scoring rules. In the context of this study, the MOs within a single CPE have been scored dichotomously using an automatic rule-based scoring system. Alternatively dichotomously scored MOs could be grouped systematically to form polytomous score units (PSU). The distribution of conditional covariance among the MOs within a CPE could then be compared across the different scoring methods when contextual and scoring dependencies, to varying degrees, exist.

If a scaling model is selected that under-represents the latent space in terms of non-nuisance factors, some portion of the residual covariance is likely due to dimensionality within the assessment. As the underlying correlation between traits changes, the magnitude of the residual covariance will also be affected. In the context of this study, different scaling methods could be implemented when the dimensional structure is known to assess the magnitude of residual covariances produced for each scaling method.

When CPEs are present in a mixed format assessment, it is likely that all three sources of LID discussed above are present to some degree. Of great interest is how the magnitude and direction of the residual covariances will be affected when all three are

present in known amounts, using different combinations of scaling procedures and scoring methods.

Research Questions

The above considerations lead to three main research questions and associated sub questions listed below:

1. Does grouping the dichotomously scored measurement opportunities of a complex performance assessment into one or more polytomous score units reduce or control the amounts of residual covariance due to the typical nuisance factors (contextual/scoring factors) associate with this item type?
    - 1.1. Is polytomous scoring equally effective in reducing/controlling the amount of residual covariance as the amount of context and scoring dependency increases?
    - 1.2. Is the effect consistent over various sample sizes?
    - 1.3. Is the effect consistent over various test lengths?
    - 1.4. Is the effect consistent when difference proportions of CPE items are included in the test?
2. What are the effects of using different IRT scaling procedures on unidimensional and multidimensional data as it relates to the amount and direction of residual covariance between items?
    - 2.1. How does the correlation between the two traits, if present, impact the amount of residual covariance?
    - 2.2. Is the effect consistent over various sample sizes?

2.3. Is the effect consistent over various test lengths?

2.4. Is the effect consistent when difference proportions of CPE items are included in the test?

3. How do the three sources of LID discussed above interact and manifest themselves in the residual covariance structures across different combinations of scoring methods and scaling procedures employed in this study?

3.1. Are these effects consistent over various sample sizes?

3.2. Are these effects consistent over various test lengths?

3.3. Are these effects consistent when difference proportions of CPE items are included in the test?

CHAPTER III

METHODS

In this study a computer simulation was used to addresses the research questions presented in Chapter Two. Simulation studies allow any number of experimental conditions that may not be readily observable in real testing situations to be set and carefully controlled. In addition, simulation allows for easy and consistent replication of the conditions, which would be prohibitively expensive in a study with live subjects. While it is acknowledged that a simulation of testing scenarios will never accurately characterize the true complexity of real data (and hence not allow for definitive conclusions), simulations are useful for framing general patterns and trends of a limited selection of phenomena of interest.

Conditions of Study

For this study two sets of conditions will be used to generate data: test format variables, and variables manipulating the amount of underlying item dependencies. A third set of conditions, a set of selected scoring and scaling methods, will then be applied to the generated data sets. A description of conditions used to generate response data for the simulation follows. All levels of all conditions are fully crossed, and each resulting combination of conditions are replicated 50 times. Table 1 displays the design of the simulation data generation.

Table 1: Simulation Conditions

Sample Size	Number of Items	Percent CPE	Scoring and Contextual Dependencies											
			None				Low				High			
			Association Between Traits											
			1.0	0.7	0.5	0.2	1.0	0.7	0.5	0.2	1.0	0.7	0.5	0.2
1000	120	30	50	50	50	50	50	50	50	50	50	50	50	50
		50	50	50	50	50	50	50	50	50	50	50	50	50
3000	60	30	50	50	50	50	50	50	50	50	50	50	50	50
		50	50	50	50	50	50	50	50	50	50	50	50	50

Test Format Variables

Sample Sizes

Two samples sizes were selected to represent the upper and lower ends of the typical number of test-takers that might be administered a single form of a large-scale mixed format test containing CPEs. The larger sample size of 3,000 represents a sample size that is on the larger end of the sample size spectrum. A sample of size of 1,00 was selected to represents the smaller end of this spectrum. The number of test-takers is near the minimum recommended sample size for the model selected for scaling.

Test Lengths

Test lengths of 60 and 120 items were also simulated. While the actual numbers of items appearing on a high volume standardized assessment would vary greatly, these

two numbers of items were selected to represent what would be considered an average length test and a long test in a large-testing program.

Proportion of Complex Performance Assessments Items

The last test format condition that is manipulated in this study is the proportion of CPE items that are included in each simulated test form. Two proportions of CPE items, 30% and 50% of the total of items, were set. Because of the increased cost of development and increased seat-time that is incurred by the use of CPEs, it is unlikely that a large scale mixed format test would use a majority of these items. The 50% level is an upper limit to the number of these items than would be included. A test that that consists of 30% CPEs is more representative of what would be expected in a mixed format test.

Local Item Dependency Conditions

Complex statistical dependencies between items may arise within and between CPE measurement opportunities. This study is concerned with dependencies that arise from three sources: underlying dimensionality, context dependencies, and scoring dependencies. The section below explains the conditions that manipulated the amount of LID within and between CPE items due to these sources

Underlying Dimensionality

Two general dimensionality structures were considered in this study: (1) all items are characterized by a single dimension, and (2) CPE items are represented with one dimension and SRSR items represented by a second. Within the two-dimensional structure, the level of association between the two traits will also be manipulated. These levels of associations were set to $\rho_{\theta_M \theta_C} = 0.2$, $\rho_{\theta_M \theta_C} = 0.5$, and $\rho_{\theta_M \theta_C} = 0.7$. It can be assumed that the two components of a single exam are likely to be at least moderately correlated and uncorrelated traits are not likely to be observed and practice. For this reason, a condition where CPE and SR traits are not correlated was not considered in this study. Coupling the three two-dimensional structures with the unidimensional condition yields a total of four levels of this condition.

Dependency due to Context/Scoring

The amounts of dependences due to the other two factors will be considered together. Three levels of this condition are set for this study: no dependencies, low amounts of context/scoring dependencies, and high amounts of context/scoring dependencies. Under the high condition, responses are related as much to scoring and contextual factors as they are to the specified primary abilities. The low condition was specified by creating parameters for data generation that were half that of the parameters used for generating data at the high dependencies condition.

Data Generation

Data for all of the replications of all conditions was generated using the computer program, MIRTGEN (Luecht, 2004), which uses a 3pl compensatory multidimensional IRT model (see the brief example in chapter 2) to generate dichotomous responses a specified number of items. MIRTGEN allows up to 50 latent dimensions in the data generation process. A vector of means and standard deviation for N dimensions, as well as the lower diagonal of the correlation matrix defining the association between each pair of latent traits, is specified in the command file, and a vector of abilities is generated for each simulated test-taker from a multivariate normal distribution. This study assumes that all latent factors/traits are distributed as a multivariate standard normal distribution, with all correlations between dimensions set to zero, with the exception of the association between the first two traits. The number and nature of the traits used in the data generation process are described in the following section and in Table 2.

MIRTGEN also requires a set of item parameters to be specified in a separate file. The item parameter file must include a vector of discrimination parameters, a, for each latent dimension, a location or difficulty parameter, d, for each item, and optionally, a lower asymptote parameter, c, for each item. The following process was used to create the "true" item parameters used in the data generation process:

1. A vector of a parameters (one for each dimension specified in the data generation process) was sampled from a log-normal distribution. The number and nature of the *a* parameters are discussed in detail with reference to Figure 1 in the next section.

2. The location parameters were sampled from a standard normal distribution. Extreme values (sampled parameters where |d|>3) were discarded and replaced with new sampled values. This distribution of location parameters simulates items that are located across the entire ability scale.
3. For SRSR items, c parameters were sampled from a uniform distribution with lower and upper bounds of 0.1 and 0.25 respectively. As the c parameter is often described as a pseudo-guessing parameter, these values are consistent with the values that might be expected in a selected response item with 4 to 5 response options.
4. CPE items are assumed in this study to consist of largely open ended responses, where guessing behaviors would not significantly impact the probability of a correct response. The c parameters were set to zero for each of these items.

For each replication of each condition, the generating parameters and true ability scores were retained. An example of both the command language and the required item parameter file is found in appendix A.

Structure of the Generated Response Data

Figure 1 (in a SEM path diagram format) illustrates how the dimensional structure was manipulated in the data generation process to form locally dependent item sets due to underlying dimensionality, context and scoring. At the top of figure 1, two ovals (labeled SRSR and CPE) represent two primary latent traits—one that quantifies the knowledge, skills, and abilities measured by the SR items and one that one that quantifies the

knowledge, skills, and abilities measured by the CPE MOs. In the unidimensional study condition, these two latent factors would be replaced with a single latent factor. The curved arrow between these two traits indicates the traits are associated with one of the correlations levels tested in the study.

Below the two primary latent traits are two distinct sets of boxes that represent the responses to the SR items (labeled with an M) and the CPE MOs (labeled with an S). The singled headed arrow linking each item to its respective latent trait is a factor loading that describes the strength of the relationship between the item and trait. Each of these factor loadings, in the context of IRT, represents the set of true a parameters that are used in the data generation process. The values of the true a parameters are sampled from a lognormal distribution with a mean of zero and a variance of 0.1. This yields a set of parameters with an expected mean of one and typical values ranging between 0.8 and 1.2. Extreme values ($a>1.5$ and $a<0.5$) were discarded and replaced with resampled values from the same distribution.

In addition to the two primary latent ability factors, Figure 1 also shows a number of secondary latent factors that are associated with the CPE MOs. Two of these latent factors are labeled "Context 1" and "Context 2". For each simulated test, half of the CPE MOs are associated with the first context factor and the second half of CPE MOs associated with the second context factor. This roughly simulates a testing scenario where every test form includes two large CPE items. Each CPE would contain multiple scored tasks (represented by the individual MOs in figure 1) and would share a common contextual setting (e.g. a common topic, common reading passages, a set of common

direction or resources, etc.). The data is generated with respect to the context factors, but is systematically left out of all scaling procedures. This induces contextual associations between the MOs loading to the same context factor, by under-representing the latent space. The magnitude of the created dependencies is directly related to the magnitude of the IRT a parameters used to generate the data.

In this study, three levels of non-dimensional dependencies (i.e. context and scoring dependencies) were simulated. In the no dependency case, the factor loading between context factors and item would be set to zero, indicating context factors do not influence the responses of test-takers. In the high context/scoring dependency, the true parameters used for data generation were sampled from a lognormal distribution with a mean of -0.5 and a standard deviation of 0.16, producing a set of a parameters with an expected mean of 0.6 and a typical range of parameters with values between 0.4 to 0.8. In the low context/scoring dependency, the true parameters used for data generation were sampled from a lognormal distribution with a mean of -1.2 and a standard deviation of 0.32, producing a set of a parameters with an expected mean of 0.3 (half that of the high condition) and a typical range of parameters with values between 0.1 to 0.5.

Scoring factors were simulated in a similar fashion to the context factors. Here, within each context linked CPE, small clusters of MOs are assumed to be linked by a set of scoring rules that may explicitly lead to dependencies among responses. As with the context factor, simulated responses are generated with this factor structure in place, but scaled with this factor systematically left unspecified, inducing association between the MOs linked to a common scoring factor. Figure 1 shows that for every three MOs within

a single CPE there is a separate scoring factor specific to only those three MOs. Like above, the loadings between each MO and its scoring factor represent the values of the IRT a used to generate the response data.

In the no dependency case, the factor loading between scoring factors and item would be set to zero, indicating that either MOs are each independently evaluating or the scoring rules are designed in such a way that they do not induce a relationship between responses. In the high context/scoring dependency condition, the *a* parameters used for data generation were sampled from a lognormal distribution with a mean of zero and a standard deviation of 0.1. This produced a set of *a* parameters with an expected mean of one and a typical range between 0.8 to 1.2. In the low context/scoring dependency, the true parameters used for data generation were sampled from a lognormal distribution with a mean of -0.7 and a standard deviation of 0.2, producing a set of a parameters with an expected mean of 0.5 (half that of the high condition) and a typical range of parameters with values between 0.3 to 0.7. Here, the high condition was initially chosen to produce parameters as discriminating as the primary trait parameters, as the scoring rules that create item chains and share scored objects are likely to produce large dependencies. The low condition was designed to produce parameters that are roughly half that of the high dependency condition.

Table 3 provides a summary of the total number of latent dimensions simulated to generate the response data across the conditions of the study. Table 4 displays the pattern of a parameters that would be used to generate data for a 60 item test, with 30% of the test consisting of CPE items, with two underlying dimensions, when scoring and context

dependencies are present. In this table, a value of one indicates that an a parameter would be sampled from the specified lognormal distribution, and a zero indicates there is no expected relationship between the response and the latent factor. An example of an entire item parameter file used in data generation is found in Appendix A.

Table 2: Number of Latent Dimensions used to Generate Data

Total Items	Percent CPE items	Underlying Dimensionality	Context Scoring Factors	**Number of Factors**			
				Primary	Context	Scoring	Total
60	30	1D	No	1	0	0	1
			Yes	1	2	6	9
		2D	No	2	0	0	2
			Yes	2	2	6	10
	50	1D	No	1	0	0	1
			Yes	1	2	10	13
		2D	No	2	0	0	2
			Yes	2	2	10	14
120	30	1D	No	1	0	0	1
			Yes	1	2	12	15
		2D	No	2	0	0	2
			Yes	2	2	12	16
	50	1D	No	1	0	0	1
			Yes	1	2	20	23
		2D	no	2	0	0	2
			yes	2	2	20	24

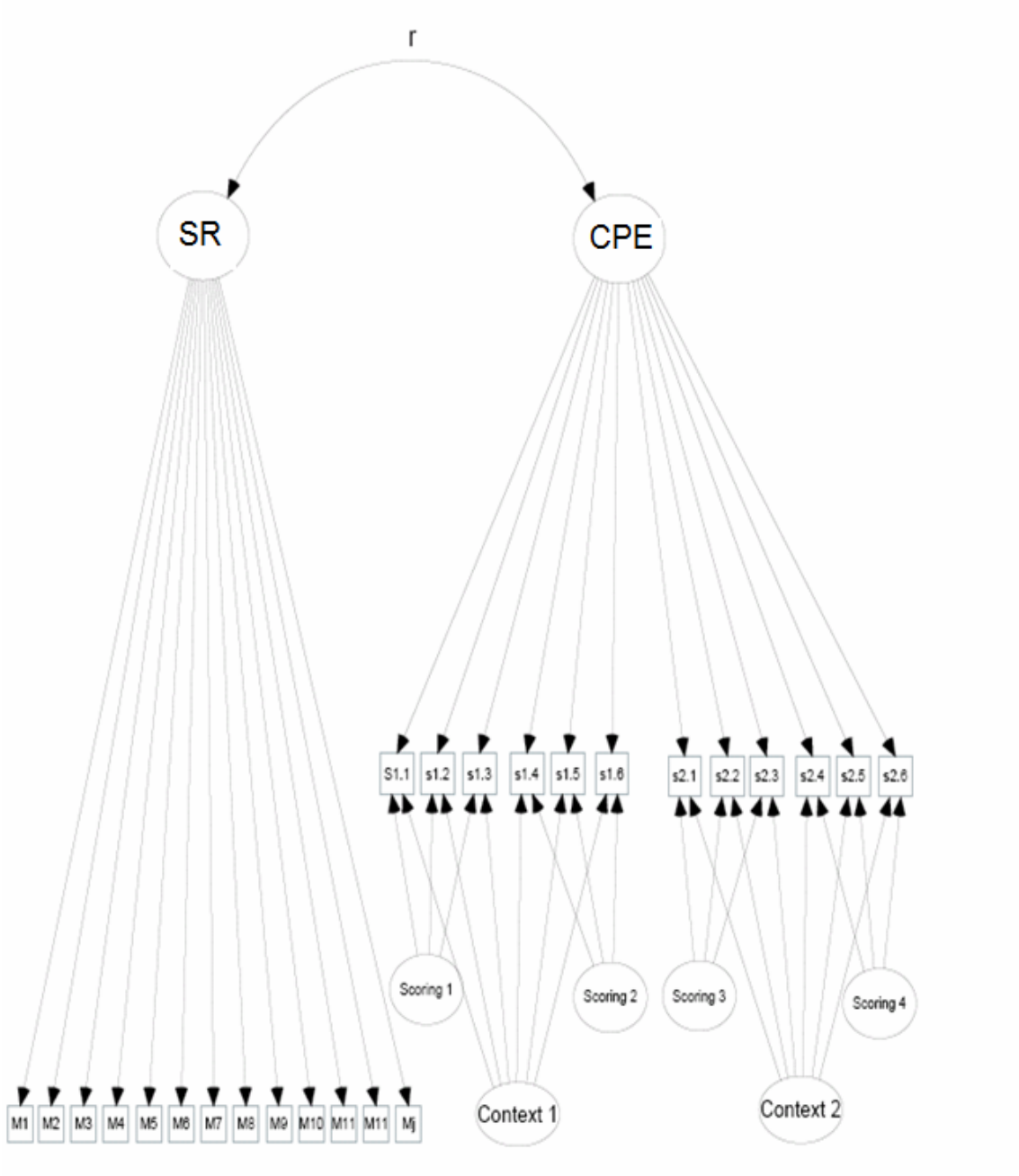

Figure 1: Schematic Diagram of the Structure of Generated Data

Table 3: Pattern Matrix for IRT a Parameters

Item	a_{SR}	a_{CPE}	a_{con1}	a_{con2}	a_{sco1}	a_{sco2}	a_{scor3}	a_{sco4}	a_{sco5}	a_{sco6}
SR1	1	0	0	0	0	0	0	0	0	0
SR2	1	0	0	0	0	0	0	0	0	0
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
SR42	1	0	0	0	0	0	0	0	0	0
CPE1	0	1	1	0	1	0	0	0	0	0
CPE2	0	1	1	0	1	0	0	0	0	0
CPE3	0	1	1	0	1	0	0	0	0	0
CPE4	0	1	1	0	0	1	0	0	0	0
CPE5	0	1	1	0	0	1	0	0	0	0
CPE6	0	1	1	0	0	1	0	0	0	0
CPE7	0	1	1	0	0	0	1	0	0	0
CPE8	0	1	1	0	0	0	1	0	0	0
CPE9	0	1	1	0	0	0	1	0	0	0
CPE10	0	1	0	1	0	0	0	1	0	0
CPE11	0	1	0	1	0	0	0	1	0	0
CPE12	0	1	0	1	0	0	0	1	0	0
CPE13	0	1	0	1	0	0	0	0	1	0
CPE14	0	1	0	1	0	0	0	0	1	0
CPE15	0	1	0	1	0	0	0	0	1	0
CPE16	0	1	0	1	0	0	0	0	0	1
CPE17	0	1	0	1	0	0	0	0	0	1
CPE18	0	1	0	1	0	0	0	0	0	1

Scoring and Scaling Procedures

The primary purpose of this study is to investigate the effect that different methods of producing scores for CPEs and scaling those resulting scores have on the residual covariances among items. The intent is to illustrate and explore the practical effects that these scoring and/or scaling methods would have when applied under different situations where one would expect to find item dependencies due one or more of

the three sources of interest. The next section will describe the scoring and scaling methods as they are applied in the study

Scoring Procedures

The data generation process described above produces dichotomous scores for both SR and CPE items under different dimensional conditions and with varying degrees of dependencies among items due to context or scoring. As mentioned in Chapter Two, the most typical methods for dealing with dichotomously scored items that are assumed to be related due to scoring and or context, is to create one or more polytomous items from the related dichotomous responses. In this study polytomous score units were created from the generated dichotomous responses by summing the dichotomous responses for the clusters of MOs that are related to a common scoring factor (as illustrated in Figure 1) for each simulated test-taker. As each scoring cluster was composed of three MOs, each polytomous item that results from this collapsing would have four score categories, with scores ranging from 0-3. Table 4 displays the number of items present across test format condition under the different scoring methods. This method was applied to each of the replications across each condition, thus producing both a dichotomously and polytomous scored version of the same raw data. This allows for a direct comparison of the residual covariance across the two scoring methods.

This approach to creating polytomous scores still allows for the various polytomous items formed within a complex performance assessment to be related to some degree through the existence of a common context factor. The formation of a

single polytomous score from all the measurement opportunities in a CPE may reduce contextual associations as well, and would be the most favorable condition. In this study the large number of MOs within a single CPE would have produced polytomous items with too many categories to use commercially available software in the scaling process, and so was not employed.

Table 4: Number of CPE Items

Total Items	Percent CPE	Scoring Method		
		SR	CPE	
			Dichotomous	Polytomous
120	30	84	36	12
	50	60	60	20
60	30	42	18	6
	50	30	30	10

Scaling Procedures

Creating polytomous scores from a set of related items is one solution for dealing with residual covariance created due to extraneous scoring and contextual factors, but cannot explain correlated item residuals due the presence of more than one primary trait. Three IRT based approaches to scaling that employ different assumptions about the underlying dimensional structure of the data are used in this study to explore the magnitude of residual covariance of CPE items when the scaling method is aligned or not aligned with the known underlying structure. The scaling methods used in this study were selected to demonstrate some very different, but operationally feasible, approaches

that could be taken into account for residual covariance. In theory, any number of scaling methods or models could be employed. The three methods are described below. All IRT calibrations and production of latent scale scores were completed using the IRT Control Language (ICL) computer program (Hanson, 2003). For all replications, the estimated parameters and latent abilities were retained.

Single Trait Approach

The first method employs a strictly unidimensional view of the data, regardless of the known underlying structure. Under this method, item parameters for the SR questions are estimated first, using a 3pl IRT model. The SR item parameters are then used as an anchoring block in a second calibration that includes the CPEs items. The result of fixing the values of the SR parameters in this calibration is that the CPE item parameters are constrained to be on the same scale as the SR items. A 2pl IRT model is applied to the CPE items when the data is dichotomously scored CPEs, and the Generalized Partial Credit Model is applied to the polytomously scored CPEs. Using this single set of item parameters, a latent trait score is estimated for each test-taker using an Expected A Posteriori (EAP) scoring algorithm.

The latent trait estimates produced in this method represent a strictly unidimensional view of the data, as it is assumed that the CPE items measure the same thing as the SR items. When the data are truly unidimensional, it is expected that any residual covariance present will be attributable to context and scoring association among the items on a single CPE. In the two dimensional condition, it is expected that sizable

amounts of residual covariance between the CPE MOs will be observed as the IRT scaling model that does not account for the second primary trait. As the correlation between traits decreases, the magnitude of residual covariance is expected to increase.

Composite Trait Approach

The second scaling methods also produces a single ability estimate based on the entire set of items. All items are calibrated together using the appropriate IRT model (i.e., the 3pl for the SR items, the 2pl for dichotomous CPEs, and the GPCM for polytomous simulations). These item parameters are then used to produce EAP latent trait scores. This type of calibration results in a composite unidimensional trait, where the contribution of the individual items to the score estimate is weighted by their relative influence or sensitivity to the underlying trait being estimated.

If the underlying dimensional structure of the data is unidimensional, these scaled scores should be nearly identical to those produced under the first scaling method. Little residual covariance due to dimensionality is expected to be observed. If the structure is two-dimensional, the resulting scale is a reference composite that represents, at least in part, both dimensions. Less residual covariance is expected in all of the two dimensional conditions under this approach when compared to the first scaling method. Additionally, as the representation of both types of items, CPE and SR, equalize (i.e. the 50% CPE items condition), it is expected that the composite trait will better represent both scales, and the amount of residual covariance will decrease in comparison to the 30% CPE item condition.

Separate Scales Approach

The last scaling procedure assumes an underlying two dimensional structure. In this scaling method, the two sections of the examination are scaled separately using the appropriate IRT models[2]. Each separate set of item parameters is then used to produce two separate EAP scores, one for each section of the test. It is expected that this method will always account for the entire latent space (outside of what is attributable to scoring and context factors) regardless of the underlying dimensionality or association between traits, and hence will result in negligible amounts of residual covariance.

Interaction of Conditions

In this study all conditions were fully crossed with both the scaling and scoring methods described above. This leads to a very rich simulation where the various sources of LID are completely confounded and allows for an exploration of the magnitude of residual covariance when different combinations of scaling methods and scoring methods are applied to data with varying degrees of dimensionality and context/scoring dependencies.

[2] In essence, this condition represents a Principled Multidimensional structure of the underlying test (i.e. items were simulated under conditions of initial simple structure, as if this was the test developers intent). The feasibility of developing such a test in practice is beyond the scope of this research.

Calculation of Residual Covariance

For each replication, a standardized residual covariance structure, via Yen's Q_3(Yen, 1984, 1993), was created for the six combination of scaling and scoring methods using the following steps.

1. For each simulated test-taker, an expected score was calculated for each item using the estimated latent abilities and item parameters from a scoring/scaling procedure (see equations 2.1-2.5 for an description of this process[3]).
2. These expected scores were then subtracted from the corresponding raw data to create a set of residuals for each test-taker on each item.
3. A correlation matrix of all item-pair residuals was created—this is a Q_3 matrix.
4. The expected mean of the Q_3 distribution (see equation 2.15[4]) was subtracted from each residual correlation to center the expected distribution at zero. This aids in the interpretation of both magnitude and direction of the residual covariance.

One advantage in using Q_3 in LID studies is the ability to decompose the correlation matrix into subsets of items-pairs that are expected to have non-zero residual covariance. The magnitude of the residual correlations for created subsets can be

[3] In the single trait and composite trait scaling methods a single set of item parameters and latent abilities is used to create a single set of expected scores. The separate scales method require that two set of expected scores be created—one from calibration/scaling of the SR items and one from the calibration/scaling of the CPA MOs. The two resulting set of expected scores were then joined to perform step 3.

[4] For the composite trait and single trait scaling methods, the total number of items *J*, is the total number of item on the test. For the separate scales methods, *J* represents the number of items on the two subtests.

aggregated to allows comparisons of the amount of evidence of LID in each subset (Y. W. Lee, 2004). For this study each residual correlation matrix was decomposed into two sets: Correlations of all MOs within a single CPE and correlations of the MOs across the two CPEs on a given form. A third section , the correlations between MOs and SR items., was not considered as the simulation employed simple structure, thus the expected residual covariance between the items types is zero. Figure 2 presents this decomposition graphically.

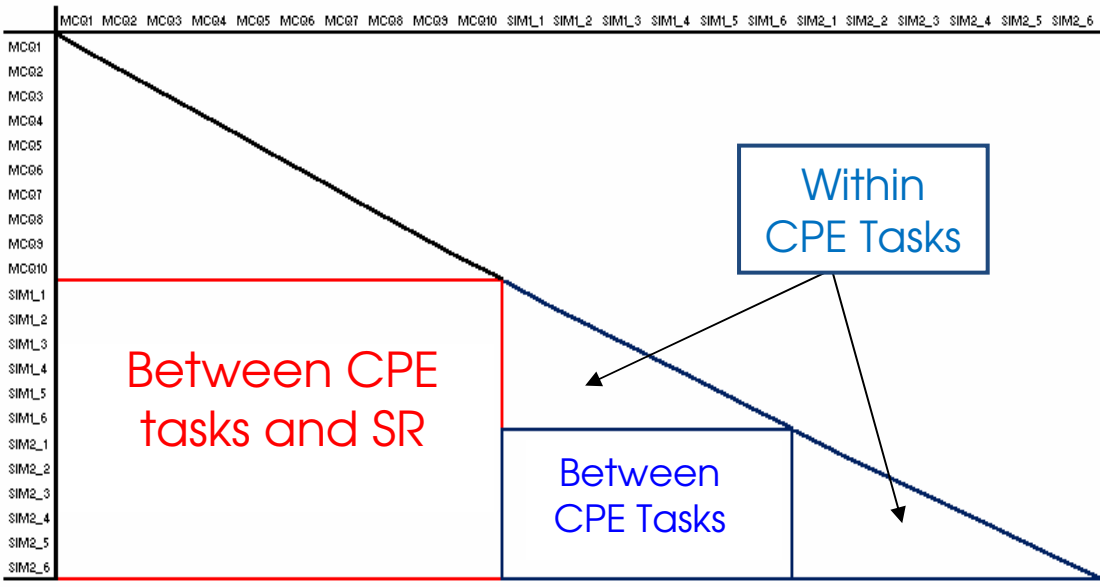

Figure 2: Partitioning of the Residual Covariance Matrix for Analysis

Measurement Opportunities within a single CPE (indicated as "Within CPE Tasks" in Figure 2) have the potential of producing convergent residual covariance due to shared scoring and scaling factors. If the CPEs represent a second trait as well, this also could result in convergent residual covariance (evidence that these MOs share a common

unspecified trait). The MOs from opposite CPEs (indicated as "Between CPE Tasks" in Figure 2) share no common contextual or scoring factors. Any observed residual covariance in this group of correlations is expected to be convergent and attributable to multidimensionality within the test. In each replication for each combination of conditions, the mean, standard deviation , minimum and maximum, as the range of residual correlation is retained for each of the two separate partitions of the Q_3 matrix. These statistics will be used to compare both the average amount and distributions of the residual covariances under the employed scoring and scaling procedures across all conditions.

Expected Results

Unidimensional Structure with no Context/Scoring Dependencies

This combination of conditions represents a situation where both the assumption of local independence and unidimensionality are strictly met. The resulting residual covariance can be viewed as a baseline. All methods of scoring and scaling should result in negligible amounts of residual covariance for all subsets of the Q_3, across all test condition variables.

Unidimensional Data with Contextual/Scoring Dependencies

When only contextual and scoring dependencies are present, residual covariance can only be reasonable expected in the within CPE MOs subset of the Q_3 matrix. As the

scoring and contextual factors become more influential (as represented by the higher *a* parameters used for generation in the high dependency condition) , more residual covariance is expected. Applying polytomous scoring should be effective in reducing the average amount of residual covariance in the within MOs correlations and/or eliminating the extreme dependencies. This effect is expected to be consistent across all three scaling methods and all test format conditions.

Two-dimensional Data with No Contextual/Scoring Dependencies

When the data is two dimensional and no context or scoring factors are present, similar amounts of positive residual covariance are expected in the between and within MOs subsets of the Q_3 matrix for the single trait and composite trait scaling conditions. As the correlation between traits increases, the amount of residual covariance for both of these methods is expected to decrease. As the percent of CPE included increases, the composite trait should produce less residual covariance than the single trait approach. The separate scale approach is expected to reduce levels of residual covariance to negligible levels for all three subsets of the Q_3 matrix, regardless of the correlation between traits or the percent of CPEs MOs included. The effects should be consistent across all scoring methods and all levels of the test-length and sample size conditions.

Two-dimensional Data with Contextual/Scoring Dependencies

Of all comparisons, this set of conditions is of the most interest. When both contextual/scoring factors are present and the data is two-dimensional, the two separate

scales condition with polytomous scoring should produce negligible amounts of residual covariance in all subsets of each Q_3, as it deals with all sources of simulated LID. This effect is expected to be consistent across all test format conditions, dimensionality conditions, and the high/low context and scoring dependency conditions.

It is conceivable that the sources of contextual and dimensional covariance will interact in unpredictable ways in the other two scaling methods. As the correlation between traits decreases, the magnitude of residual covariance will increase for all subsets of the Q_3 matrix. Further, polytomous scoring should be effective in reducing the overall amount of residual covariance and/or eliminating the extreme dependencies caused by scoring factors. It is less clear how the levels of contextual/scoring dependencies tested in this study will interact with the various dimensional structure, as the inclusion of many ancillary factors can drastically affect the representation of the CPE trait that is projected onto the SR trait (single trait scaling condition) or the reference composite trait (the composite trait scaling condition).

CHAPTER IV

RESULTS

The results of this study are presented in four sections. Section one presents the results of the baseline (unidimensional data with no nuisance dependencies) analysis, where the residual covariance across all test format conditions is expected to be zero. The last three sections address the three research questions of the study separately. The second section examines the effects that adopting a polytomous scoring method has on the magnitude and distribution of the residual correlations of tasks within the same CPE when contextual/scoring dependency are present. Section three explores how the residual covariance structures of CPE tasks differ under three different scaling method as the dimensional structure changes. The last section explores how the residual covariance structures differ when various combinations of scoring and scaling are applied to the data.

Baseline Analysis

The baseline analysis examines the magnitude of the residual correlation when the data is simulated to be unidimensional and with no additional nuisance factors contributing to the responses of candidates. In these conditions, regardless of sample size, test length, or percent of CPE, zero covariance is expected. Further, as no context/scoring factor comes into play, there is no distinction between the two partitions (within and between CPEs) of the Q_3 matrices. The intent of this analysis is to simply

verify that the Q_3 statistics performs as described in the literature[5] and thus allow a basis of comparison for the analyses that follow.

Table 5 shows some descriptive statistics for the Q_3 correlations of the pertinent conditions. For each replication, the mean Q_3 value, the standard deviation of the all of the Q_3, and the range of the Q_3 values were recorded. The table contains the average Q_3 value, aggregated over the 50 replications within each combination of the conditions. As expected, the "Within CPE" and "Between CPE" partitions of the Q_3 matrix are nearly identical in terms of these statistics within each set of conditions. Average standard deviations across conditions also tend to be consistent and small across all conditions, with a slight decrease in variation observed as sample size increases. The larger ranges of Q_3 tend to be associated with dichotomous scoring and longer tests, and are likely due to the simple fact that, given more items[6], the chance of observing a large value of Q_3 by chance increases. Mean values of Q_3 tend to be close to zero in all conditions, with means ranging from -0.03 to 0.01. Figure 3 provides a closer look at how these mean Q_3 values change across the various conditions condition

[5] Recall that each Q_3 matrix was corrected subtracting of the expected value of Q_3 as described by Yen (1984,1993). This means that under conditions where there should be no residual covariance, the expectation is that mean value of Q_3 should be zero after adding in the correction factor –(1/J-1)

[6] In comparison to the dichotomous scoring, the creation of polytomous items reduces the total number of items by 1/3

Table 5: Baseline Residual Covariance Statistics

Sample	Items	Percent	Scoring	Scaling	Range		Mean		SD	
					Between	Within	Between	Within	Between	Within
1,000	60	30	Dichotomous	One	0.16	0.16	0.00	0.00	0.03	0.03
				Composite	0.16	0.16	0.00	0.00	0.03	0.03
				Separate	0.18	0.17	0.01	0.01	0.04	0.04
			Polytomous	One	0.10	0.08	-0.03	-0.03	0.03	0.03
				Composite	0.10	0.08	-0.03	-0.03	0.03	0.03
				Separate	0.11	0.11	0.00	0.01	0.04	0.04
		50	Dichotomous	One	0.19	0.18	0.00	0.00	0.03	0.03
				Composite	0.19	0.18	0.00	0.00	0.03	0.03
				Separate	0.19	0.19	0.00	0.01	0.03	0.03
			Polytomous	One	0.12	0.13	-0.02	-0.02	0.03	0.03
				Composite	0.12	0.13	-0.02	-0.02	0.03	0.03
				Separate	0.13	0.13	0.01	0.01	0.03	0.03
	120	30	Dichotomous	One	0.20	0.20	0.00	0.00	0.03	0.03
				Composite	0.20	0.20	0.00	0.00	0.03	0.03
				Separate	0.20	0.20	0.00	0.00	0.03	0.03
			Polytomous	One	0.14	0.14	-0.01	-0.01	0.03	0.03
				Composite	0.14	0.14	-0.01	-0.01	0.03	0.03
				Separate	0.14	0.14	0.01	0.01	0.03	0.03
		50	Dichotomous	One	0.21	0.21	0.00	0.00	0.03	0.03
				Composite	0.21	0.21	0.00	0.00	0.03	0.03
				Separate	0.21	0.21	0.00	0.00	0.03	0.03
			Polytomous	One	0.16	0.17	-0.01	-0.01	0.03	0.03
				Composite	0.16	0.17	-0.01	0.00	0.03	0.03
				Separate	0.16	0.17	0.01	0.01	0.03	0.03
3,000	60	30	Dichotomous	One	0.10	0.09	0.00	0.00	0.02	0.02
				Composite	0.10	0.10	0.00	0.00	0.02	0.02
				Separate	0.11	0.12	0.01	0.01	0.02	0.02
			Polytomous	One	0.06	0.05	-0.03	-0.03	0.02	0.02
				Composite	0.06	0.05	-0.03	-0.03	0.02	0.02
				Separate	0.09	0.08	0.01	0.01	0.03	0.03
		50	Dichotomous	One	0.11	0.11	0.00	0.00	0.02	0.02
				Composite	0.11	0.11	0.00	0.00	0.02	0.02
				Separate	0.12	0.11	0.00	0.00	0.02	0.02
			Polytomous	One	0.08	0.07	-0.03	-0.02	0.02	0.02
				Composite	0.08	0.07	-0.03	-0.02	0.02	0.02
				Separate	0.08	0.08	0.01	0.01	0.02	0.02
	120	30	Dichotomous	One	0.12	0.11	0.00	0.00	0.02	0.02
				Composite	0.12	0.11	0.00	0.00	0.02	0.02
				Separate	0.12	0.12	0.00	0.00	0.02	0.02
			Polytomous	One	0.08	0.08	-0.01	-0.01	0.02	0.02
				Composite	0.08	0.08	-0.01	-0.01	0.02	0.02
				Separate	0.09	0.09	0.01	0.01	0.02	0.02
		50	Dichotomous	One	0.13	0.12	0.00	0.00	0.02	0.02
				Composite	0.13	0.12	0.00	0.00	0.02	0.02
				Separate	0.13	0.12	0.00	0.00	0.02	0.02
			Polytomous	One	0.10	0.10	-0.01	-0.01	0.02	0.02
				Composite	0.10	0.10	0.00	-0.01	0.02	0.02
				Separate	0.10	0.10	0.01	0.01	0.02	0.02

Figure 3 displays the average Q_3 values further aggregated over the "Within CPE" and "Between CPE" distinctions. Each panel in the figure represents one combination of the levels of the test length, sample size, and percent CPE study conditions. Across the x-axis are the three scaling methods tested in this study. The mean values of Q_3 are plotted on the y-axis with a square for dichotomous scoring and a diamond for polytomous scoring. These figures demonstrate that while all the values are close to the expected value of zero, there are some important and systemic differences. The largest deviations from the expected values of zero come when the CPEs are scored polytomously and these items are scaled jointly with the dichotomous SRs. In each of these cases, the average values of Q_3 are smaller than the expected values of Q_3. This deviation is particularity large when the test length is small and the percent of CPEs is large. In contrast, the same scaling methods applied to dichotomously scored CPEs produce mean residual correlations that are perfectly aligned with expected value of zero. The separate scaling approach produces slight positive deviations for both scoring types, though the deviations are larger for polytomously scored CPEs.

These results indicate that when all of the items scaled together have the same number of score categories (i.e. all items are dichotomous or polytomous items with the same number of score categories), the average amount of residual covariance according to the correct Q_3 matrix is at or near zero as expected. For tests that contain mixed formats of questions (i.e. dichotomous and polytomous items), the correction of subtracting the expected mean value Q_3 proposed by Yen (1984, 1993), is successful only as the number of items increase and the proportion of the various item types are equal.

As this correction does appear to be asymptotically appropriate, the corrected Q_3 will be used for the remainder of the study. However, it is suggested that further investigation into the expected value of the distribution of Q_3 when there is no LID should be investigated for mixed format tests.

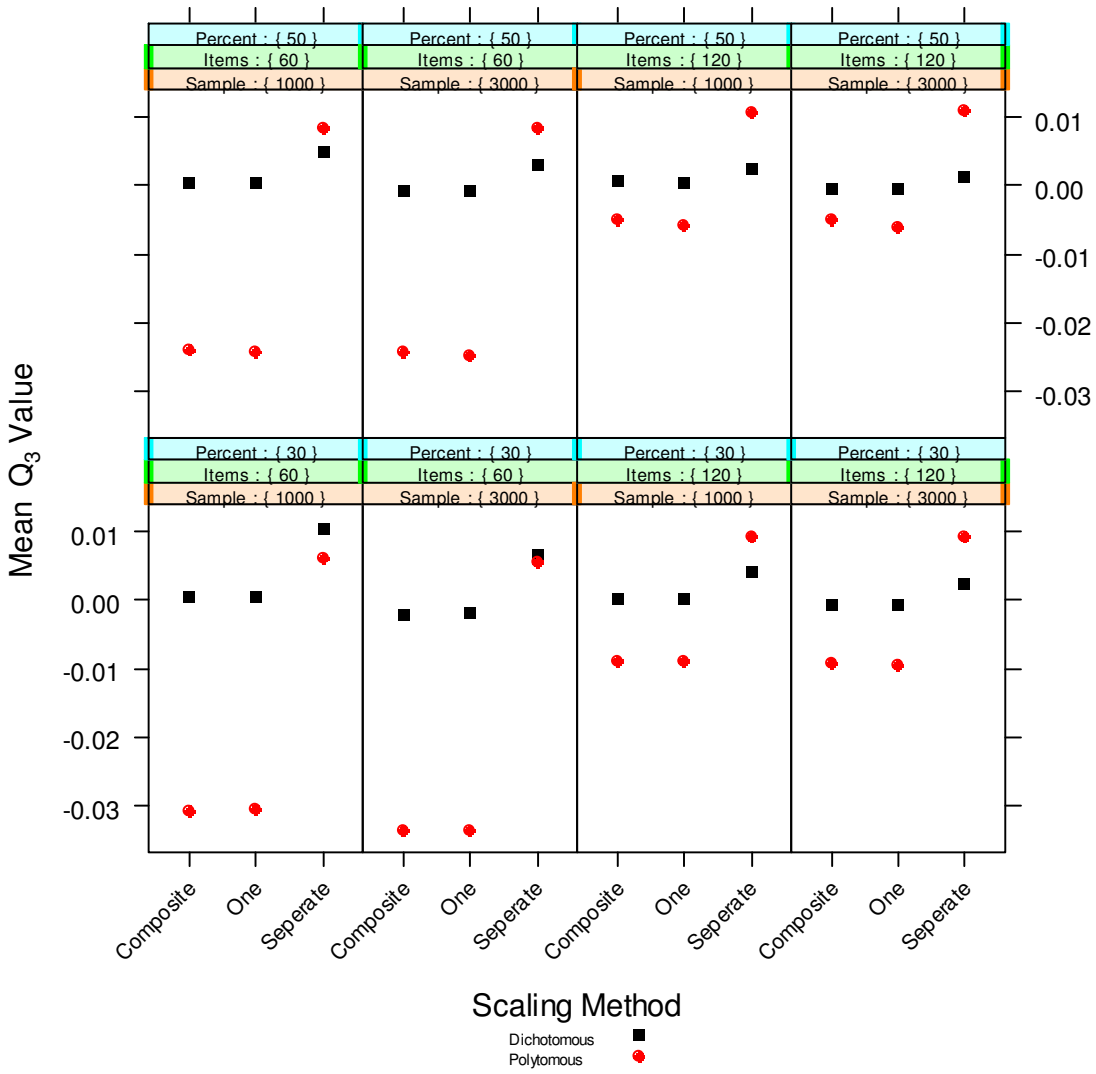

Figure 3: Mean Residual Covariance for CPE Items for Baseline Condition

Scoring Method Comparisons

The most popular solution for dealing with contextual and/or scoring dependencies among sets of items is to create polytomous items from the individually scored dichotomous items. This set of results, which addresses the first research question of this study, asks if this approach is effective in a setting where the MOs of a CPE share a common contextual setting and scoring rules that create association among the some of the MOs. It is noted here that in this study only associations due to scoring will be ameliorated by forming polytomous units, as the MOs from a shared scoring factor were used to form polytomous units. The simulated contextual factor may still cause association among the responses of the newly formed polytomous items, thus allowing some residual covariance between these items. In light of this limitation, dichotomously scored MOs are affected by both scoring factors and contextual factors, thus allowing a comparison of the two scoring types, as polytomous scoring should resolve the LID due to scoring.

The results presented below represent only the simulated data that have LID due to nuisance factors and not dimensionality. Further, as the influence of these nuisance factors were systematically simulated only to affected MOs within the same CPE, the data used in the presentations below are limited to only include the "Within CPE" subset of the Q_3 matrix. Last, the data are aggregated over the three different scaling methods, as the differences across the methods are consistent over all dependency and test format conditions when the underlying dimensional structure of the data is unidimensional (see the results tables in Appendix D).

Table 6 displays the mean, standard deviation and range of the "Within CPE" portions of the Q_3 matrices of the 50 replications of each condition aggregated across the three scaling methods. Overall, the average range and standard deviation decrease as the simulated level of dependencies decrease. As the number of items increase, the range tends to increase and the standard deviation decrease. Sample size and the percent of CPE items has little affect on the average range and standard deviation. There is a marked difference in both of these measures of spread when comparing across scoring method. The Q_3 correlations produced under polytomous scoring have much lower ranges and standard deviations than the Q_3 correlations produced under dichotomous scoring. This is true for both the high and low contextual/scoring dependencies conditions.

Table 6 also shows that across conditions where contextual LID is present, the mean Q_3 values are higher for polytomous scoring when compared to dichotomously scored CPEs. The difference between polytomous and dichotomous mean Q_3 increases as the level of simulated LID increases[7]. There appears to be no changes in mean Q_3 for either scoring methods as sample size changes. Increases in test length and percent of CPE items include lead to only a minor decrease in mean Q_3, and these changes are consistent across scoring methods.

[7] As mentioned above, polytomous scoring in the context of this study does not deal with the context factors, and hence will still cause some level of association between polytomously scored items. It is unclear how the mean residual correlations would compare if a scoring approach that dealt contextual factors as well had been employed in this study.

Table 6: Comparison CPE Residual Covariance for Unidimensional Data Under Different Scoring Methods

Sample Size	Item Length	Level of Dependency	Scoring Method	30% CPE			50% CPE		
				Range	Mean	SD	Range	Mean	SD
1000	60	High	Dichotomous	0.44	0.12	0.12	0.46	0.08	0.10
			Polytomous	0.10	0.16	0.04	0.15	0.14	0.04
		Low	Dichotomous	0.27	0.05	0.06	0.28	0.03	0.05
			Polytomous	0.10	0.07	0.04	0.13	0.06	0.03
		None	Dichotomous	0.16	0.00	0.03	0.19	0.00	0.03
			Polytomous	0.09	-0.02	0.03	0.13	-0.01	0.03
	120	High	Dichotomous	0.46	0.08	0.09	0.49	0.07	0.07
			Polytomous	0.15	0.15	0.04	0.18	0.14	0.04
		Low	Dichotomous	0.30	0.04	0.05	0.31	0.03	0.04
			Polytomous	0.15	0.08	0.04	0.18	0.07	0.04
		None	Dichotomous	0.20	0.00	0.03	0.21	0.00	0.03
			Polytomous	0.14	0.00	0.03	0.17	0.00	0.03
3000	60	High	Dichotomous	0.39	0.12	0.12	0.41	0.08	0.10
			Polytomous	0.07	0.16	0.03	0.11	0.15	0.03
		Low	Dichotomous	0.22	0.04	0.05	0.23	0.03	0.04
			Polytomous	0.08	0.07	0.03	0.09	0.06	0.02
		None	Dichotomous	0.10	0.00	0.02	0.11	0.00	0.02
			Polytomous	0.06	-0.02	0.02	0.08	-0.01	0.02
	120	High	Dichotomous	0.41	0.08	0.09	0.42	0.07	0.07
			Polytomous	0.11	0.15	0.03	0.12	0.14	0.03
		Low	Dichotomous	0.24	0.03	0.04	0.25	0.03	0.03
			Polytomous	0.10	0.08	0.03	0.12	0.07	0.02
		None	Dichotomous	0.12	0.00	0.02	0.12	0.00	0.02
			Polytomous	0.08	0.00	0.02	0.1	0.00	0.02

Figure 4 displays the mean Q3, aggregated across the two sample size conditions, with the scoring methods within a set of conditions side by side for a direct comparison. Each panel represents one combination of the levels of the test length and percent CPE variables. Across the x-axis are the three levels of simulated contextual/scoring

dependencies. The two bars in each panel represent the mean residual correlation values for the two scoring conditions. The trend noted in the description of Table 6 is more clearly observed here. In all panels, the mean amount of residual covariance is higher for polytomous scoring than dichotomous scoring of CPEs when high or low levels of context/scoring dependencies are present. The panels also show small decreases in residual correlation across both scoring types as the number of items increase and as the percent of CPE items increase.

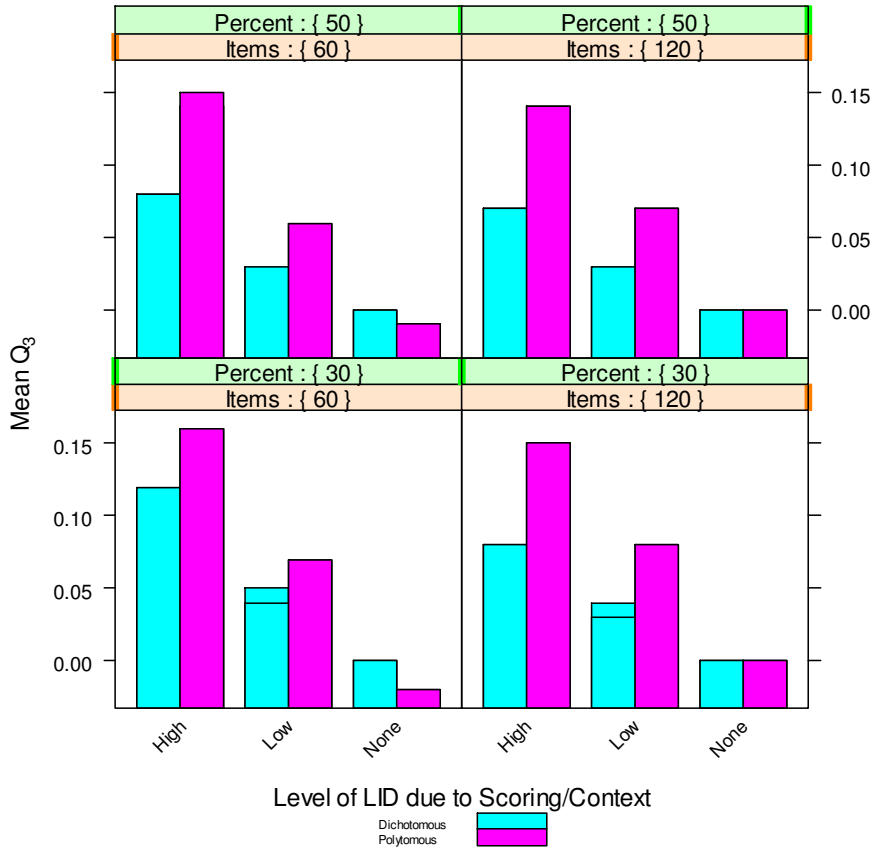

Figure 4: Comparison of Average Q_3 Values Under Polytomous and Dichotomous Scoring Over Three Levels of Nuisance Dependency

Figure 5 allows for the examination of the whole distribution of Q_3 values rather than just the mean. Each row of panels in this figure corresponds to one of the two levels of simulated contextual/scoring dependencies. The panels in each row are the four combinations of test length and percent CPE conditions. The two boxplots in each panel represent the distribution of Q_3 values for the different scoring methods. The results in Table 6 indicate that there is little difference in the distribution of Q_3 across sample size, so the box plots in Figure 5 include values from both the 1,000 and 3,000 person sample size conditions.

While the same trend mention above is clear (polytomous scoring lead to slightly higher mean residual correlations), Figure 5 clearly shows that the distributions under dichotomous scoring have a much larger range and variance, and lead to a great number of extreme dependencies. The absence of these extreme Q_3 correlations, especially in the high context/scoring dependency condition (see the bottom row of the figure), suggests that polytomous scoring is successful in eliminating the extreme instance of LID due to scoring factors.

These results suggest that if a test can safely be considered unidimensional, and there is reason to believe that scores on CPE MOs are associated due to scoring considerations, polytomous scoring may be employed to reduce the number of extreme dependencies, which could directly impact the quality of the estimated item parameters of the associated items. This study does not support the claim that polytomous scoring can reduce the overall amount of residual covariance. However, as mentioned above, the process of forming polytomous items in this study did not address the existence of

contextual dependencies, which would certainly contribute to the mean amount of residual covariance in the polytomously scored CPEs. Further research is needed to test whether polytomous scoring of CPE tasks could also be used to eliminate LID due to shared contextual factors, and if this would reduce the average amount of residual covariance to that of dichotomously scored CPEs that share context and scoring factors.

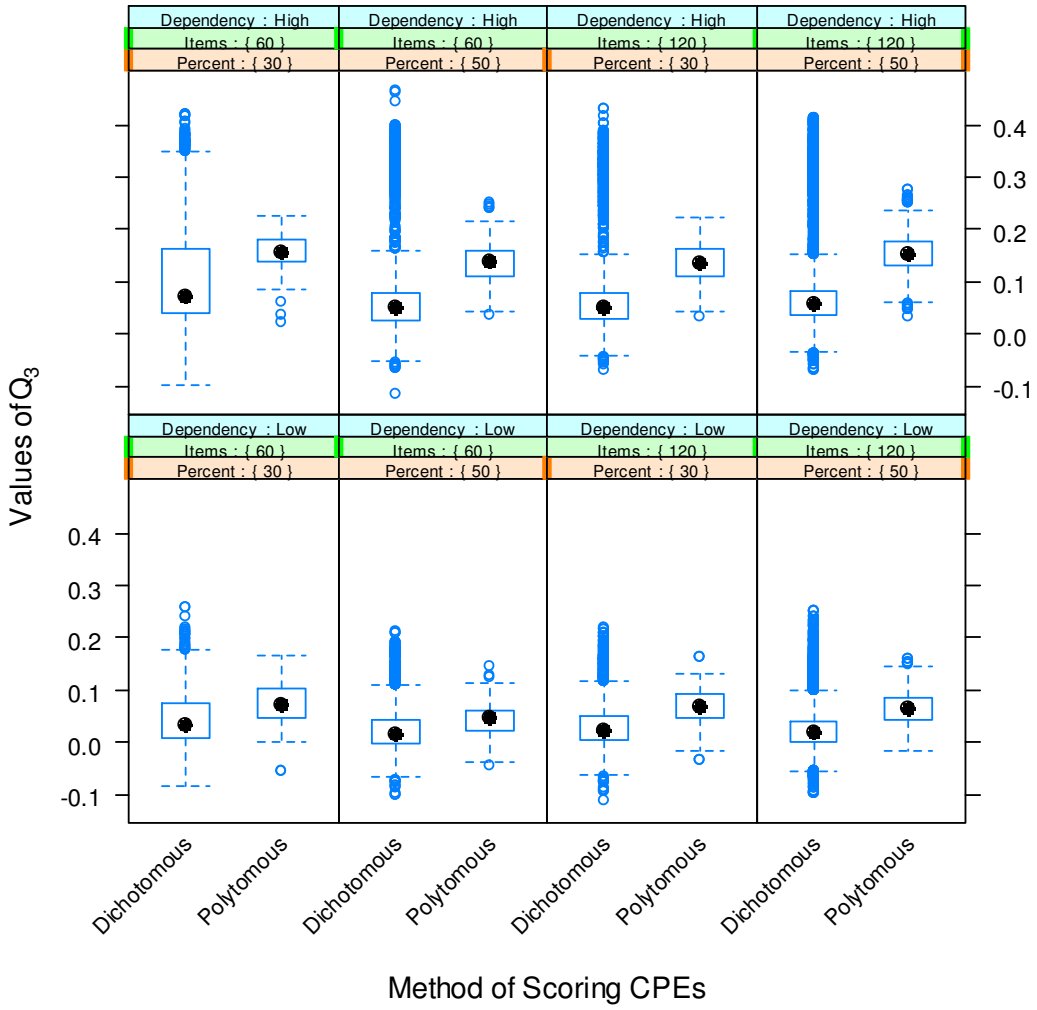

Figure 5: Distribution of Q3 under Dichotomous and Polytomous Scoring

Scaling Method Comparisons

This section of the results examines the magnitude of residual covariance when the data has unidimensional or two-dimensional structure (with varying levels of correlation between the two traits) and no LID due to nuisance factors under three different, but operationally plausible, scaling methods. Since residual covariance is a direct measure of how well a condition variable(s) explains the latent space and therefore the response patterns of the test-takers, the scaling methods that produce the least residual covariance for a given dimensional structure best meet the assumptions of local independence. This analysis uses the subset of simulated data where no contextual or scoring factors influence the response. Also, as only the set of data replications with no contextual LID are considered, and polytomous scoring is used only to ameliorate LID of this nature, only the dichotomously scored CPEs need be considered here. In addition, the absence of LID due to context or scoring factors renders the distinction of "Within CPEs" and "Between CPES" irrelevant[8]. The results below use the all of the inter-CPE residual correlations.

Table 7 shows some descriptive statistics for the Q_3 correlations of the pertinent conditions. For each replication, the mean Q_3 value, the standard deviation of the all of the Q_3, and the range of the Q_3 values were recorded. The table contains the average Q_3 values aggregated over the 50 replications within each combination of conditions. Looking first at the measure of spread, both measures are smaller under the separate scaling method than either the compost or single traits scaling methods if the data is two-

[8] Only context or scoring factors make the amount of residual covariance different in the within and between CPE section of the residual correlation matrix

dimensional (i.e. correlations between traits of 0.2, 0.5, or 0.7). The range and standard deviation increase as the correlation between traits decreases within any set of test format conditions for both of the joint calibration approaches, where the separate scaling method produces a stable range and standard deviation across the various correlations between traits for any given combination of test format conditions. For all scaling methods, an increase in the percent of CPE items or the number of test-takers lead to a small reduction both measures of spread. An increase in test length results in a slightly larger range of scores for only the composite or single trait methods, but the standard deviation for method was unaffected as the test length increased.

The pattern of mean values of Q_3 for the different scaling methods and across percent CPE and test length[9] is depicted is Figure 6. Across the x-axis are the various correlations between the traits (r-1.0 signifying unidimensionality). The mean value of Q_3 is plotted on the y-axis. A separate line is plotted in each panel representing the how the mean amount of residual covariance changes as the correlation between traits increases for each scaling condition. Across all conditions, the separate scaling method produced a mean residual covariance that is at most very close to zero and never greater than 0.01. This is clearly seen in Figure 6 as the line with points plotted as triangles running parallel to the bottom of each panel. The other two scaling methods result in, on average, large amounts of residual covariance when the two traits have a small correlation. As the correlation between traits approaches one, the mean residual covariance approaches zero. Both joint scaling approaches show marked improvement

[9] An examination of table 7 shows means values are consistent across sample size, so it is not included in the figure.

when the percent of CPE items increase to 50 percent, with the composite trait more drastically impacted than the single trait method of scaling.

This is evidence that the separate scaling approach, when there are no nuisance sources of LID, produces overall amounts of residual covariance that are close to zero regardless of the correlation between traits. This is not surprising as only CPE items are used to create the CPE ability score and these items were generated to constitute a single trait.

Table 7: Comparison of Residual Covariance Under Different Scaling Methods

Sample Size	Test Length	Correlation Between Traits	Scaling Method	30% CPE			50% CPE		
				Range	Mean	SD	Range	Mean	SD
1000	60	0.2	One	0.27	0.28	0.06	0.24	0.13	0.04
			Composite	0.26	0.27	0.06	0.25	0.04	0.04
			Separate	0.18	0.01	0.04	0.19	0.01	0.03
		0.5	One	0.23	0.20	0.05	0.21	0.09	0.04
			Composite	0.23	0.20	0.05	0.22	0.06	0.04
			Separate	0.18	0.01	0.04	0.19	0.01	0.03
		0.7	One	0.19	0.12	0.04	0.19	0.05	0.03
			Composite	0.19	0.12	0.04	0.19	0.04	0.03
			Separate	0.18	0.01	0.04	0.19	0.01	0.03
		1	One	0.16	0.00	0.03	0.19	0.00	0.03
			Composite	0.16	0.00	0.03	0.19	0.00	0.03
			Separate	0.17	0.01	0.04	0.19	0.01	0.03
	120	0.2	One	0.32	0.28	0.06	0.27	0.10	0.04
			Composite	0.31	0.26	0.06	0.29	0.05	0.04
			Separate	0.20	0.00	0.03	0.22	0.00	0.03
		0.5	One	0.28	0.19	0.05	0.24	0.08	0.04
			Composite	0.28	0.19	0.05	0.24	0.05	0.04
			Separate	0.20	0.00	0.03	0.22	0.00	0.03
		0.7	One	0.24	0.12	0.04	0.23	0.05	0.04
			Composite	0.24	0.12	0.04	0.22	0.04	0.03
			Separate	0.20	0.00	0.03	0.22	0.00	0.03
		1	One	0.20	0.00	0.03	0.21	0.00	0.03
			Composite	0.20	0.00	0.03	0.21	0.00	0.03
			Separate	0.20	0.00	0.03	0.21	0.00	0.03
3000	60	0.2	One	0.24	0.28	0.05	0.20	0.12	0.04
			Composite	0.23	0.26	0.05	0.20	0.05	0.04
			Separate	0.12	0.01	0.03	0.12	0.00	0.02
		0.5	One	0.20	0.20	0.04	0.14	0.08	0.03
			Composite	0.20	0.20	0.04	0.15	0.05	0.03
			Separate	0.12	0.01	0.03	0.12	0.00	0.02
		0.7	One	0.15	0.12	0.03	0.12	0.05	0.02
			Composite	0.15	0.12	0.03	0.12	0.04	0.02
			Separate	0.12	0.01	0.03	0.12	0.00	0.02
		1	One	0.10	0.00	0.02	0.11	0.00	0.02
			Composite	0.10	0.00	0.02	0.11	0.00	0.02
			Separate	0.11	0.01	0.02	0.11	0.00	0.02
	120	0.2	One	0.29	0.27	0.05	0.21	0.09	0.03
			Composite	0.28	0.25	0.05	0.23	0.05	0.04
			Separate	0.12	0.00	0.02	0.13	0.00	0.02
		0.5	One	0.25	0.19	0.05	0.16	0.07	0.03
			Composite	0.26	0.19	0.05	0.16	0.05	0.02
			Separate	0.12	0.00	0.02	0.13	0.00	0.02
		0.7	One	0.18	0.12	0.03	0.14	0.05	0.02
			Composite	0.18	0.11	0.03	0.14	0.04	0.02
			Separate	0.12	0.00	0.02	0.13	0.00	0.02
		1	One	0.11	0.00	0.02	0.13	0.00	0.02
			Composite	0.11	0.00	0.02	0.13	0.00	0.02
			Separate	0.12	0.00	0.02	0.13	0.00	0.02

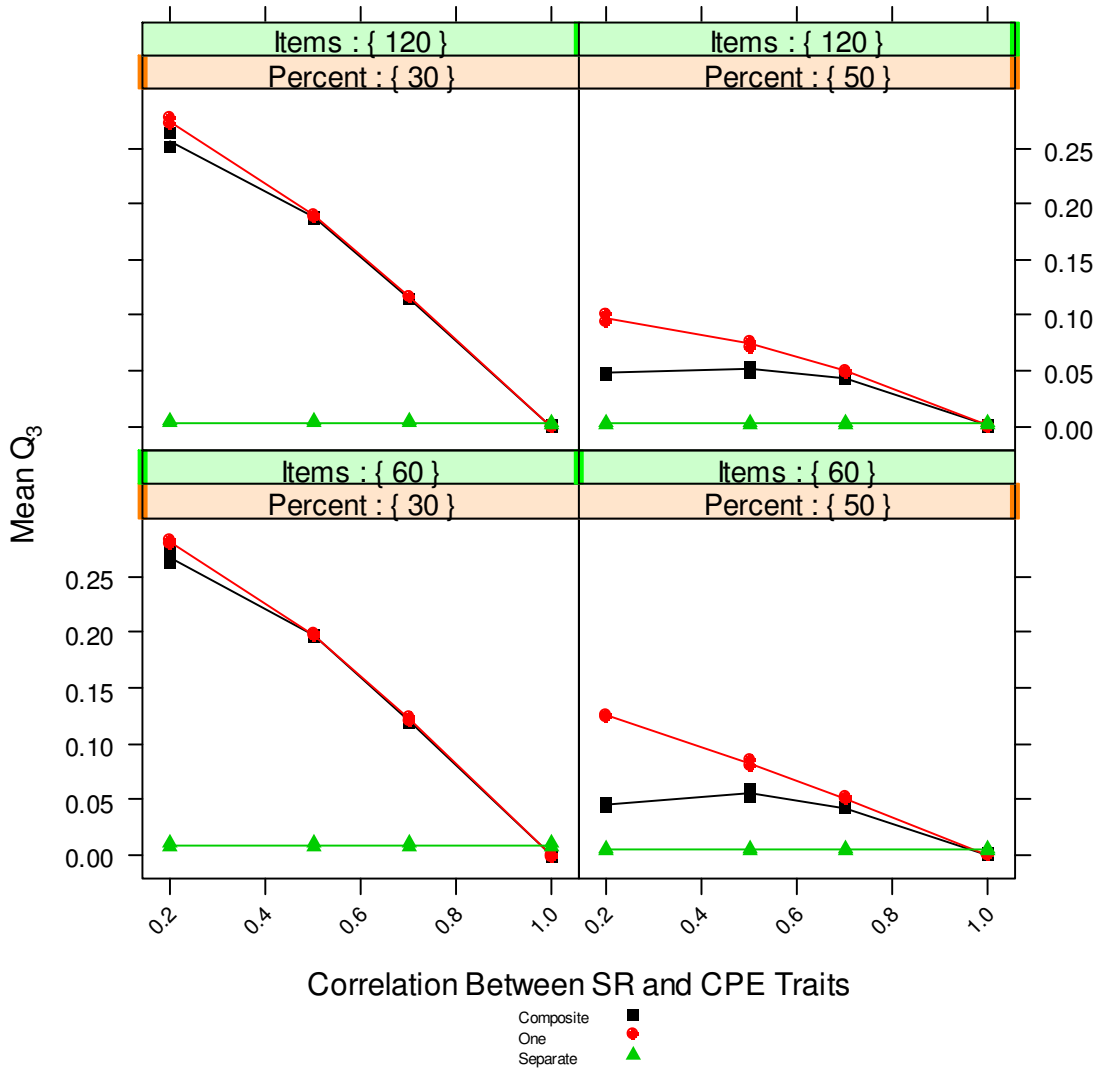

Figure 6: Comparison of Average Q_3 Values of Scaling Methods over Four Dimensional Structures

Scaling and Scoring Combinations

Up to this point, the analyses deal with situation where one source (nuisance or dimensionality) is responsible for the residual covariance, and a scaling or scoring choice

can be made to deal appropriately with the latent space and issues of LID. When CPE are included in a mixed format test, it is likely that both nuisance factors such as context or scoring and underlying dimensionality will be present to some degree. The results in this situation examine the residual correlations under different scaling and scoring combinations in situations where the four different dimensional structures in this study are coupled with the three different levels of contextual and scoring. Here, both subsets of the Q_3 matrices described in Chapter, "Within CPEs" and "Between CPEs", are considered separately, as the amount of residual covariance will not be the same in these sections as nuisance factor will only contribute residual covariance to the "Within CPEs" section of a given Q_3 matrix.

Tables 8 and 9 contain descriptive statistics for Q_3 in the 30% CPE and 50% CPE respectively, with both sets of results aggregated over sample size and test length[10]. Overall, the patterns of average residual covariance and measures of spread average as the percent of CPE increases are consistent with the result discussed previously. The mean values of Q_3, especially for the one trait and composite trait scaling approaches, tend to decrease (i.e. the mean values in table 9 tend to lower than the corresponding value sin table 8). The measure of spreads are consistent across the two table. These tables will be discussed together as the results tend to show the same general patterns.

[10] Sample size and test length had little effect on the average amount of residual covariance across the over conditions in the study. The standard deviation and range changed across sample sizes in ways consistent to what is reported in the results above—range tends to increase as items added and is unaffected by changes in sample size. The standard deviation is slightly smaller in the larger sample size conditions, but is unaffected by the test length. A full set of results can be found broken down by samples size and test length combination in Appendix C

The difference between the "Within" and "Between" mean correlations for each set of conditions is listed in the last column. When no contextual/scoring dependencies are simulated, the mean "Between" and "Within" residual correlations are equal. The difference between the mean "Within " and "Between" sections of the Q_3 matrix decrease as the level of contextual dependency decreases. This is expected as the only difference in residual covariance across these two subsets should be dependencies due to context or scoring, as both are equally affected by the underlying dimensionality. In addition, whenever contextual or scoring dependencies are present, the "Within CPE" Q_3 have much larger range of values and a higher standard deviation. For any given level of simulated contextual dependency ("High" or "Low"), the differences tend to increase slightly as the correlation between traits increases.

Tables 8 and 9 also allow a comparison of the scaling methods employed in this study. When the SR and CPE items are treated as distinct and scaled separately, the average amount of residual covariance, range, and standard deviation all remain the same across all simulated dimensional conditions. In the "Between CPE" portion of the residual correlation matrix under this same scaling method, the mean Q_3 value is near zero when no contextual or scoring dependencies are present, but becomes increasingly negative as the level of contextual LID increases. This slight divergent covariance between MOs from different CPEs is not evidence that the two CPEs on a single test form constitute different triats, but rather indicates that these two sets of MOs are related to other factors the not accounted for in the modeled latent space (in this case the two distinct context factors associated with the two separate CPEs).

The two joint scaling approaches (scaling all items as one trait or all items into a composite traits) shows a distinctly different pattern than that described above. These scaling methods show an increase in the mean amount of residual covariance as the correlation between traits decreases. Interestingly, as the amount of contextual/scoring LID decrease from high amounts to none, the mean Q_3 in the "Between CPE" correlations, which are only affected by dimensionality, increase. A detailed explanation of this finding is beyond the scope of this study, but the influence of ancillary factors on to the formation of a scale is an issue that should be addressed in future research.

Table 7 and 8 also allows a comparison of the two scoring methods and their interactions with the scaling methods described above. With the exception of the "Between CPE" correlations under the separate trait approach to scaling, polytomous scoring always leads to higher average Q_3 correlations, much smaller ranges of correlations, and less variance in the observed values of Q_3. Within a given level of contextual/scoring dependency and scaling method, the difference in mean Q_3 values across the two scoring methods is similar, regardless of the correlation between traits. The differences in mean Q_3 across scoring methods becomes more pronounced as the level of dependency decreases, with the largest difference between scoring types observed when there are no contextual/scoring dependencies.

Table 8: Descriptive Statistics of Q_3 Matrices in the 30% CPE Item Condition

Level of Dependency	Correlation Between Traits	Scaling Method	Scoring Method	Between CPEs			Within CPEs			Within-Between
				Range	Mean	SD	Range	Mean	SD	
High	0.20	One	Dichotomous	0.19	0.17	0.04	0.38	0.28	0.09	*0.11*
			Polytomous	0.11	0.29	0.03	0.1	0.4	0.03	*0.11*
		Composite	Dichotomous	0.19	0.17	0.04	0.38	0.28	0.09	*0.11*
			Polytomous	0.11	0.29	0.03	0.1	0.4	0.03	*0.11*
		Separate	Dichotomous	0.16	-0.05	0.03	0.45	0.08	0.11	*0.13*
			Polytomous	0.11	-0.05	0.03	0.13	0.13	0.04	*0.18*
	0.50	One	Dichotomous	0.17	0.12	0.03	0.39	0.23	0.09	*0.11*
			Polytomous	0.1	0.24	0.03	0.1	0.35	0.03	*0.11*
		Composite	Dichotomous	0.17	0.12	0.03	0.39	0.23	0.09	*0.11*
			Polytomous	0.1	0.23	0.03	0.1	0.35	0.03	*0.12*
		Separate	Dichotomous	0.16	-0.05	0.03	0.45	0.08	0.11	*0.13*
			Polytomous	0.11	-0.05	0.03	0.12	0.13	0.04	*0.18*
	0.70	One	Dichotomous	0.15	0.07	0.03	0.4	0.19	0.09	*0.12*
			Polytomous	0.11	0.18	0.03	0.11	0.3	0.03	*0.12*
		Composite	Dichotomous	0.15	0.07	0.03	0.4	0.19	0.09	*0.12*
			Polytomous	0.11	0.17	0.03	0.11	0.3	0.03	*0.13*
		Separate	Dichotomous	0.16	-0.05	0.03	0.45	0.08	0.11	*0.13*
			Polytomous	0.11	-0.05	0.03	0.13	0.13	0.04	*0.18*
	1.00	One	Dichotomous	0.13	-0.02	0.03	0.42	0.11	0.1	*0.13*
			Polytomous	0.1	0.02	0.03	0.11	0.16	0.03	*0.14*
		Composite	Dichotomous	0.13	-0.02	0.03	0.42	0.11	0.1	*0.13*
			Polytomous	0.1	0.02	0.03	0.11	0.17	0.03	*0.15*
		Separate	Dichotomous	0.16	-0.05	0.03	0.45	0.08	0.11	*0.13*
			Polytomous	0.1	-0.04	0.03	0.12	0.13	0.04	*0.17*
Low	0.20	One	Dichotomous	0.23	0.24	0.05	0.29	0.28	0.06	*0.04*
			Polytomous	0.12	0.44	0.03	0.11	0.49	0.03	*0.05*
		Composite	Dichotomous	0.23	0.23	0.05	0.29	0.27	0.06	*0.04*
			Polytomous	0.12	0.44	0.03	0.11	0.49	0.03	*0.05*
		Separate	Dichotomous	0.16	-0.02	0.03	0.25	0.04	0.05	*0.06*
			Polytomous	0.11	-0.03	0.03	0.11	0.07	0.03	*0.10*
	0.50	One	Dichotomous	0.21	0.17	0.04	0.27	0.21	0.05	*0.04*
			Polytomous	0.13	0.36	0.04	0.12	0.42	0.04	*0.06*
		Composite	Dichotomous	0.21	0.16	0.04	0.27	0.21	0.05	*0.05*
			Polytomous	0.13	0.36	0.04	0.12	0.41	0.04	*0.05*
		Separate	Dichotomous	0.16	-0.02	0.03	0.25	0.04	0.05	*0.06*
			Polytomous	0.11	-0.03	0.03	0.11	0.07	0.03	*0.10*
	0.70	One	Dichotomous	0.17	0.1	0.03	0.25	0.15	0.05	*0.05*
			Polytomous	0.13	0.26	0.04	0.12	0.32	0.04	*0.06*
		Composite	Dichotomous	0.17	0.1	0.03	0.25	0.15	0.05	*0.05*
			Polytomous	0.13	0.26	0.04	0.13	0.32	0.04	*0.06*
		Separate	Dichotomous	0.16	-0.02	0.03	0.25	0.03	0.05	*0.05*
			Polytomous	0.11	-0.03	0.03	0.11	0.07	0.04	*0.10*
	1.00	One	Dichotomous	0.14	-0.01	0.03	0.25	0.04	0.05	*0.05*
			Polytomous	0.09	-0.02	0.03	0.11	0.07	0.03	*0.09*
		Composite	Dichotomous	0.14	-0.01	0.03	0.25	0.04	0.05	*0.05*
			Polytomous	0.09	-0.02	0.03	0.11	0.08	0.03	*0.10*
		Separate	Dichotomous	0.16	-0.02	0.03	0.25	0.04	0.05	*0.06*
			Polytomous	0.11	-0.03	0.03	0.11	0.07	0.03	*0.10*
None	0.20	One	Dichotomous	0.28	0.28	0.06	0.28	0.28	0.05	*0.00*
			Polytomous	0.12	0.55	0.03	0.12	0.55	0.04	*0.00*
		Composite	Dichotomous	0.27	0.26	0.05	0.27	0.26	0.05	*0.00*
			Polytomous	0.12	0.55	0.04	0.12	0.55	0.04	*0.00*
		Separate	Dichotomous	0.16	0.01	0.03	0.16	0.01	0.03	*0.00*
			Polytomous	0.11	0.01	0.03	0.11	0.01	0.03	*0.00*
	0.50	One	Dichotomous	0.24	0.19	0.05	0.24	0.19	0.05	*0.00*
			Polytomous	0.14	0.46	0.04	0.14	0.46	0.04	*0.00*
		Composite	Dichotomous	0.24	0.19	0.05	0.24	0.19	0.05	*0.00*
			Polytomous	0.14	0.45	0.04	0.14	0.45	0.04	*0.00*
		Separate	Dichotomous	0.16	0.01	0.03	0.15	0.01	0.03	*0.00*
			Polytomous	0.11	0.01	0.03	0.1	0.01	0.03	*0.00*
	0.70	One	Dichotomous	0.19	0.12	0.04	0.19	0.12	0.04	*0.00*
			Polytomous	0.15	0.32	0.04	0.14	0.32	0.04	*0.00*
		Composite	Dichotomous	0.19	0.12	0.04	0.19	0.12	0.04	*0.00*
			Polytomous	0.15	0.32	0.04	0.14	0.32	0.04	*0.00*
		Separate	Dichotomous	0.15	0.01	0.03	0.15	0.01	0.03	*0.00*
			Polytomous	0.11	0.01	0.03	0.1	0.01	0.03	*0.00*

Table 9: Descriptive Statistics of Q_3 Matrices in the 50% CPE Item Condition

Level of Dependency	Correlation Between Traits	Scaling Method	Scoring Method	Between CPEs			Within CPEs			Within-Between
				Range	Mean	SD	Range	Mean	SD	
High	0.20	One	Dichotomous	0.20	0.14	0.04	0.41	0.23	0.08	*0.09*
			Polytomous	0.14	0.28	0.03	0.13	0.39	0.03	*0.11*
		Composite	Dichotomous	0.20	0.09	0.03	0.42	0.19	0.08	*0.10*
			Polytomous	0.14	0.27	0.03	0.13	0.38	0.03	*0.11*
		Separate	Dichotomous	0.17	-0.05	0.03	0.45	0.06	0.09	*0.11*
			Polytomous	0.12	-0.04	0.03	0.14	0.13	0.03	*0.17*
	0.50	One	Dichotomous	0.17	0.06	0.03	0.42	0.16	0.08	*0.10*
			Polytomous	0.14	0.21	0.03	0.14	0.33	0.03	*0.12*
		Composite	Dichotomous	0.18	0.05	0.03	0.42	0.16	0.08	*0.11*
			Polytomous	0.14	0.19	0.03	0.14	0.32	0.03	*0.13*
		Separate	Dichotomous	0.17	-0.05	0.03	0.45	0.06	0.09	*0.11*
			Polytomous	0.13	-0.04	0.03	0.14	0.13	0.03	*0.17*
	0.70	One	Dichotomous	0.16	0.02	0.03	0.42	0.13	0.08	*0.11*
			Polytomous	0.13	0.13	0.03	0.14	0.27	0.03	*0.14*
		Composite	Dichotomous	0.16	0.01	0.03	0.42	0.12	0.08	*0.11*
			Polytomous	0.14	0.13	0.03	0.14	0.27	0.03	*0.14*
		Separate	Dichotomous	0.17	-0.05	0.03	0.45	0.06	0.09	*0.11*
			Polytomous	0.12	-0.04	0.03	0.14	0.13	0.03	*0.17*
	1.00	One	Dichotomous	0.16	-0.03	0.03	0.44	0.08	0.08	*0.11*
			Polytomous	0.12	0.00	0.03	0.14	0.15	0.03	*0.15*
		Composite	Dichotomous	0.16	-0.04	0.03	0.44	0.08	0.08	*0.12*
			Polytomous	0.12	0.00	0.03	0.14	0.15	0.03	*0.15*
		Separate	Dichotomous	0.17	-0.05	0.03	0.45	0.06	0.09	*0.11*
			Polytomous	0.12	-0.04	0.03	0.14	0.13	0.03	*0.17*
Low	0.20	One	Dichotomous	0.23	0.14	0.04	0.30	0.18	0.05	*0.04*
			Polytomous	0.16	0.43	0.04	0.15	0.48	0.04	*0.05*
		Composite	Dichotomous	0.24	0.05	0.04	0.30	0.10	0.05	*0.05*
			Polytomous	0.16	0.38	0.04	0.16	0.44	0.04	*0.06*
		Separate	Dichotomous	0.17	-0.02	0.03	0.27	0.03	0.04	*0.05*
			Polytomous	0.12	-0.03	0.03	0.13	0.07	0.03	*0.10*
	0.50	One	Dichotomous	0.18	0.07	0.03	0.27	0.11	0.04	*0.04*
			Polytomous	0.16	0.28	0.04	0.16	0.34	0.04	*0.06*
		Composite	Dichotomous	0.20	0.05	0.03	0.28	0.10	0.04	*0.05*
			Polytomous	0.17	0.26	0.04	0.17	0.32	0.04	*0.06*
		Separate	Dichotomous	0.17	-0.02	0.03	0.27	0.03	0.04	*0.05*
			Polytomous	0.12	-0.02	0.03	0.13	0.07	0.03	*0.09*
	0.70	One	Dichotomous	0.17	0.04	0.03	0.26	0.08	0.04	*0.04*
			Polytomous	0.15	0.15	0.03	0.16	0.23	0.04	*0.08*
		Composite	Dichotomous	0.17	0.03	0.03	0.26	0.07	0.04	*0.04*
			Polytomous	0.15	0.15	0.03	0.16	0.22	0.04	*0.07*
		Separate	Dichotomous	0.17	-0.02	0.03	0.27	0.03	0.04	*0.05*
			Polytomous	0.12	-0.03	0.03	0.13	0.07	0.03	*0.10*
	1.00	One	Dichotomous	0.16	-0.02	0.03	0.27	0.03	0.04	*0.05*
			Polytomous	0.12	-0.03	0.03	0.13	0.06	0.03	*0.09*
		Composite	Dichotomous	0.16	-0.02	0.03	0.27	0.03	0.04	*0.05*
			Polytomous	0.12	-0.03	0.03	0.13	0.06	0.03	*0.09*
		Separate	Dichotomous	0.17	-0.02	0.03	0.27	0.03	0.04	*0.05*
			Polytomous	0.12	-0.03	0.03	0.12	0.07	0.03	*0.10*
None	0.20	One	Dichotomous	0.23	0.11	0.04	0.23	0.11	0.04	*0.00*
			Polytomous	0.17	0.52	0.04	0.17	0.52	0.04	*0.00*
		Composite	Dichotomous	0.24	0.05	0.04	0.24	0.05	0.04	*0.00*
			Polytomous	0.19	0.43	0.05	0.19	0.43	0.05	*0.00*
		Separate	Dichotomous	0.16	0.00	0.03	0.16	0.00	0.03	*0.00*
			Polytomous	0.13	0.01	0.03	0.12	0.01	0.03	*0.00*
	0.50	One	Dichotomous	0.19	0.08	0.03	0.18	0.08	0.03	*0.00*
			Polytomous	0.18	0.30	0.04	0.17	0.30	0.04	*0.00*
		Composite	Dichotomous	0.19	0.05	0.03	0.19	0.05	0.03	*0.00*
			Polytomous	0.19	0.25	0.04	0.19	0.25	0.04	*0.00*
		Separate	Dichotomous	0.16	0.00	0.03	0.16	0.00	0.03	*0.00*
			Polytomous	0.12	0.01	0.03	0.12	0.01	0.03	*0.00*
	0.70	One	Dichotomous	0.17	0.05	0.03	0.17	0.05	0.03	*0.00*
			Polytomous	0.15	0.16	0.03	0.15	0.16	0.03	*0.00*
		Composite	Dichotomous	0.17	0.04	0.03	0.17	0.04	0.03	*0.00*
			Polytomous	0.15	0.15	0.03	0.15	0.15	0.03	*0.00*
		Separate	Dichotomous	0.16	0.00	0.03	0.16	0.00	0.03	*0.00*
			Polytomous	0.12	0.01	0.03	0.12	0.01	0.03	*0.00*

Figure 7 is graphical representation of the mean Q_3 values in tables 7 and 8 that provide a clearer picture of the results described above. In this figure, each row represents one of the three contextual and/or scoring dependency conditions. The two left-most columns contain results for the 30 percent CPE conditions and the two right columns represent the results when the test is composed of 50 percent CPEs. For each combination of contextual LID and percent CPE there is s panel for the "Between CPE correlations and "Within CPE" correlations. Six lines, representing all possible scoring methods by scaling method combination are plotted in each panel, with the correlation between traits on the x-axis and the mean Q_3 plotted on the y-axis.

The top row of Figure 7 displays the mean Q_3 values when no LID due to context and/or scoring is simulated. In this row, the within and between panels for a given level of percent CPE are identical. In the second and third rows, where low and high levels of LID due to context are present respectively, the within panels display the same general patterns as the between panel, with a small but constant increase in the mean Q_3 values. This small increase can be interpreted as the amount of residual covariance that is attributable to the context and/or scoring factors. It is clear that a much larger portion of the observed residual covariance is attributable a failure account for the dimensional structure.

Figure 7 shows that regardless of scoring method, if the SR items and CPE items are scaled separately, only small amounts of residual covariance due to the contextual and scoring factors, which is evident by the lines elevated above zero in the within CPE

panels, remains. As with all other scaling methods, the average amount of observed residual covariance is higher under polytomous scoring.

Looking at the one trait/composite trait scaling methods, as depicted in Figure 7, as correlation between traits moved towards one, the amount of residual covariance decreases. A comparison of the two left and two right columns (30 and 50 percent CPE conditions) demonstrates when CPE and SR are equally represented, scaling all of the items jointly into a composite strait produces residual covariance only slightly higher than that produced in the separate traits scaling method. When either a composite or single trait approach is taking scaling CPE that are polytomously scored, large amounts of residual covariance are observed. The largest amounts of residual covariance are observed when no LID due to context/scoring is present, polytomous scoring is used, and a joint scaling method is employed.

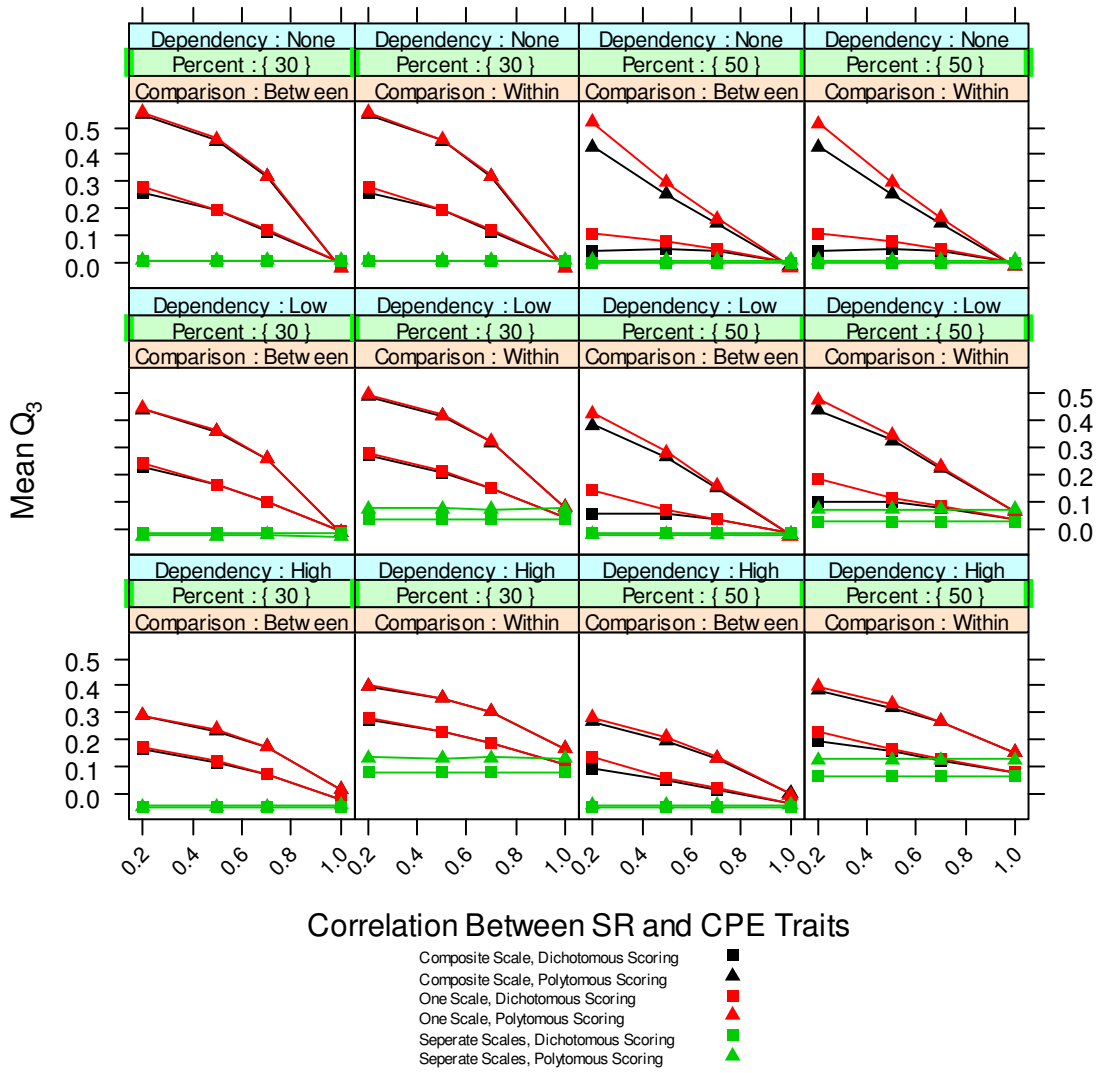

Figure 7: Mean Residual Correlation for Scoring and Scaling Method Combinations

These results indicate that if a test is known to be multidimensional, using the separate traits approach to scaling is most likely to produce residual covariances that are close to zero. Figure 7 and Tables 8 and 9 do not provide any evidence that polytomous scoring is effective in reducing the average amount of residual covariance. In fact, if

polytomous scoring applied in situations where there is little to no LID due to context or scoring factors, and a scaling method that does not address the underlying dimensionality is employed, the amount of residual covariance is likely to be extremely high. However, the average amount of residual covariance should not be the only consideration when choosing a scoring method. A large number of item pairs with large associations is equally concerning.

Figures 8 and 9 display the distributions of Q_3 for the "Within CPE" portion of the Q_3 matrices under the two most extreme scaling methods—separate scaling of the two sections or forcing the two sections to represent only one trait. The rows in both figures correspond to one of the three levels of Contextual and/or scoring dependency simulated in the study. The four panels in each row represent the four different dimensional structures simulated. Figure 8 represents the separate scales approach. It is clear that polytomous scoring, which was designed in this study to eliminate the item dependencies due to scoring factors, is very effective. In both the High and Low context/scoring LID conditions, few extreme dependencies exist after applying the polytomous scoring rules. Even when no context dependencies exist (top row of the figure), and polytomous scoring is not needed, no harm is done in terms of the overall amount of residual covariance. In spite of the slightly higher mean Q_3 under polytomous scoring[11], there are obvious advantages to employing this method.

[11] This increased mean may be due in part to this study not addressing the contextual dependencies in the contextual scoring method

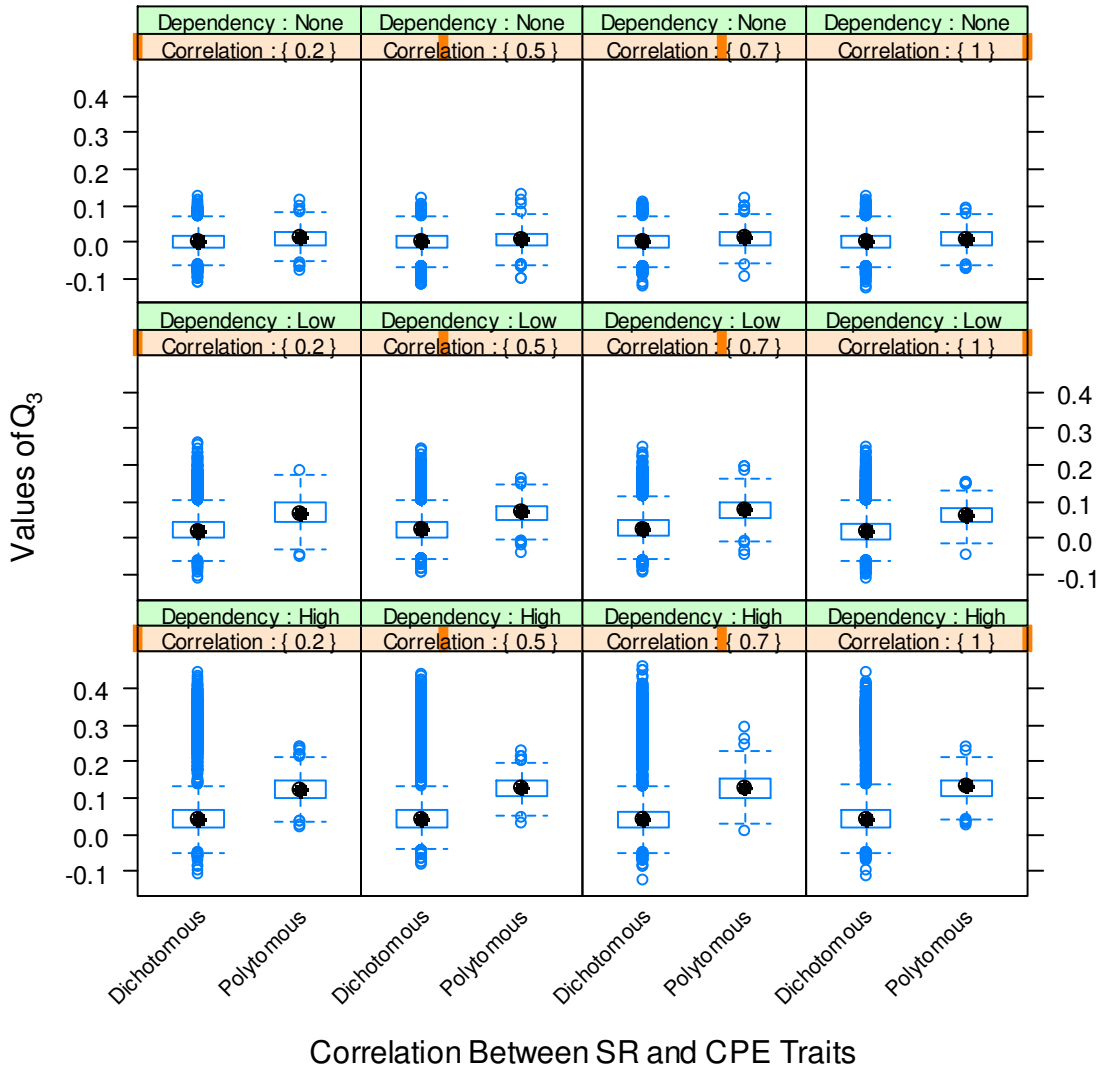

Figure 8: Distribution of Q_3 Values for Items Within the Same CPE When CPE and SR Traits are Scaled Separately

Figure 9, which displays the distributions of Q_3 when all items all anchored to the SR scale, shows a very different picture of polytomous scoring. Polytomous scoring does appear to be effective in reducing the number of extreme dependencies due to scoring

when the test is truly unidimensional (in which case the scaling method is appropriate), or the two traits are highly correlated.

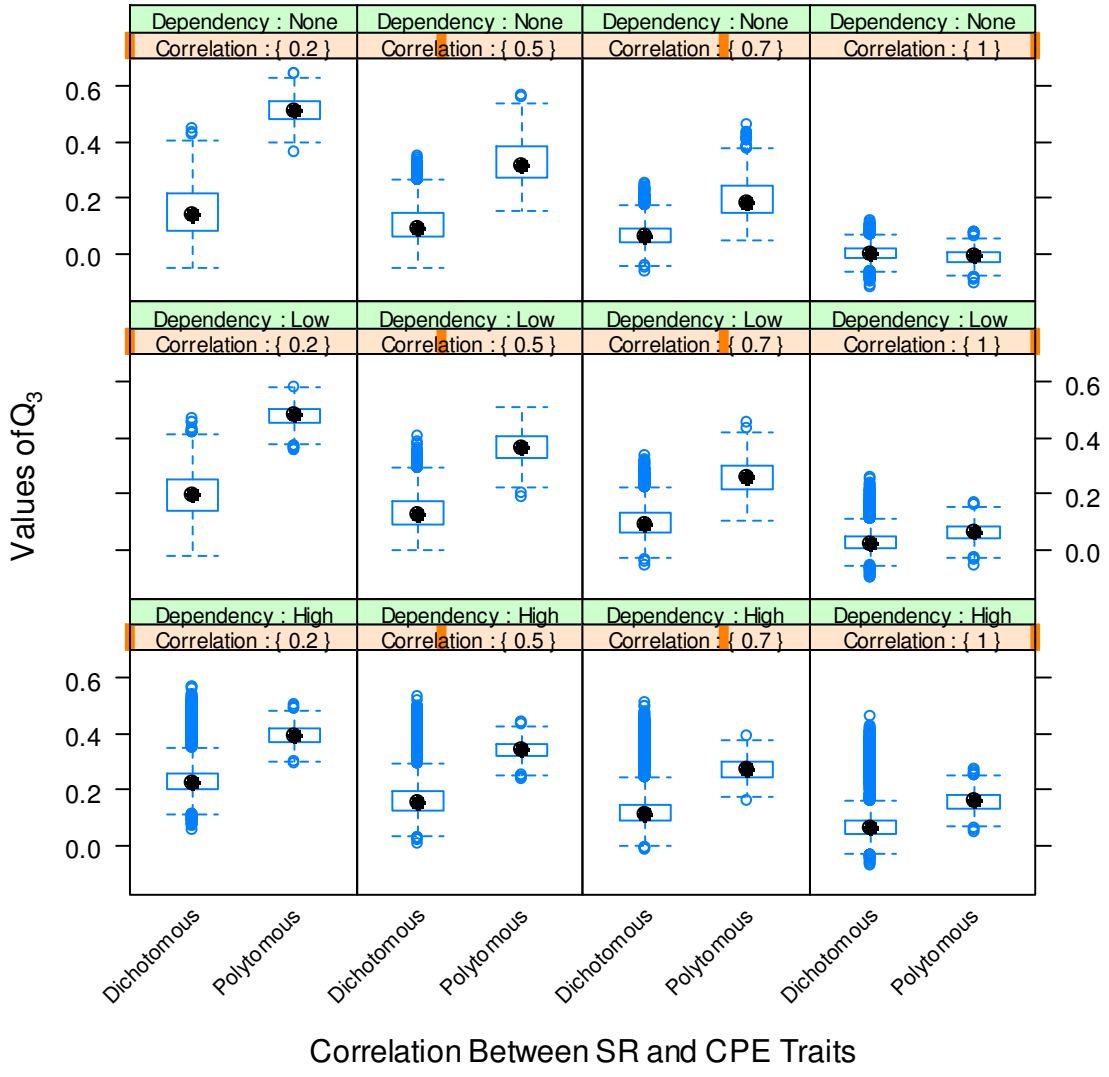

Figure 9: Distribution of Q_3 Values for Items Within the Same CPE When CPE Items are Anchored to SR Trait

CHAPTER V

CONCLUSIONS AND DISCUSSION

Summary and Implications of Findings

Research Question One asked if polytomous scoring is an effective method for dealing with CPE tasks that are related due to shared context or scoring rules that implicitly or explicitly create relationships among scored responses. The related sub questions ask if the general findings are consistent over the various test format conditions (sample size, test length, and proportion of test that were CPE items) that were included in the simulation. Overall, there was no evidence that polytomous scoring reduces the average amount of residual covariance that could be attributed to scoring factors over either of the two levels of dependency due to these sources simulated in this study. A major limitation of his study is that the process for forming polytomous items did not address LID due to the simulated contextual factors. It is possible that if polytomous score units were formed in a manner that clustered CPE tasks related to common scoring factors and contextual factors, the average amount of residual covariance could be further reduced. The study did show that polytomous scoring was effective in reducing the range and variance of Q_3 values and could eliminate the extreme dependencies that were the result of associations among CPE tasks due to scoring rules. While the high contextual dependencies clearly benefited from polytomous scoring, even the low contextual

conditions were markedly improved with polytomous scoring when the underlying data was unidimensional. The results of this part of the study were consistent over all levels of the test format variables with only small changes in the mean Q_3 as test length increased.

The second research goal was to compare the amount of residual covariance produced over four different dimensional structures when the data was scaled using three different methods. Of the three scaling methods, treating the SR and CPE items as two separate scales was consistently effective in producing Q_3 centered near zero, with small variance. This scaling method was equally successful as the correlation between traits decreased.

The other scaling approaches attempt to explain the latent space with only one latent trait score. When the data are unidimensional, these methods, like the separate scales approach, produce residual covariance structures centered at zero. If the data are two-dimensional, as the traits become more distinct, the amount of residual covariance increases. Unlike the separate trait method described above, the residual covariance structures produced change a great deal as the proportion of CPE items increases. The Composite score in particular is able to explain much more of the latent space when half of the items on the test are from each of the two underlying dimensions, as the resulting composite is influenced by the strength of the relationship of each item to its primary trait, and the number of items associated with each trait. The composite score may be an acceptable scaling choice if the correlation between traits is relatively high and the number CPE items is large. This is unfortunately not practical in operational practice.

CPE items are expensive to develop and require more seat time, and hence are typically represented in smaller numbers.

While the separate scaling option is the safest option based solely on statistical grounds, it does have some practical challenges. If two scales are to be formed, other operational and psychometric questions must be addressed. First, the process is most advantageous when the correlation between traits is moderate to low. Most tests that include both CPE and SR items may have two distinct scales, but as they are linked to a common testing purpose, the scale is likely moderately to highly correlated. As mentioned above, a composite scaling may work nearly as well. The two scales that would be formed would be based on shorter subtests, particularly the CPE section which is likely to be based on a very limited number of scored tasks. Shorter tests will in turn be less reliable and informative. Depending the use of the test, the loss of information (and hence larger standard errors and confidence intervals around the estimated abilities), may have serious implications. The creation of two scales also requires the equating of two scales, maintenance of two sets of banked item parameters, and the validation of both scales.

The test devolvement process must also be refined to ensure that each test form contains the same two scales. With CPEs, which are likely to be variable in topic and task, as well as limited in number on each form, this is no small task. Along these same lines, in large scale testing, the representation must be maintained with each administration, which would require a systematic and streamlined process of producing quality and consistent CPEs.

The last research question is based on the premise that when CPE are included in a mixed format test, both dimensionality and contextual/scoring dependencies are likely to be present. In light of this, the question asked which combination of scoring scaling would be most effective in dealing with the interaction of all of the potential sources of residual covariance. The results in the context of this study clearly indicate that the safest choice is to scale CPE and SR items separately if the test is suspected to be multidimensional. Second, coupling this approach with polytomous scoring, if context or scoring factors are believed to result in association among CPE items or tasks, can help to establish local independence within each of the subtests.

Creating a separate scale creates a smaller subtest. Collapsing dichotomous responses into polytomous items further reduces the number of items. To make this option feasible for mixed format tests that contain a limited number of CPE items, other scaling methods—using a collateral information approach (Luecht, 1993), or an item level factor analysis method (Wirth & Edwards, 2007)—may be viable options.

The results also show what does not work. If a test has both context factors and is multidimensional in nature, and a scaling method is selected that does not address the underlying dimensionality (i.e. using strictly unidimensional scaling method, as is often done in practice), the typical fix of applying polytomous scoring to help alleviate the contextual/scoring LID may do more harm than good.

The methods presented in this study should not be taken as a call for an exploratory approach to reveal the correct scoring and scaling practice. Rather, quite the opposite is advocated. Scoring and scaling decisions should be made up front, as part of

the test-development process. These methods are best used as an empirical check on preselected methods or as tool to diagnosis practical problems related to task development and performance. In a larger sense, many of the practical issues and limitations listed with the result above could be adequately dealt with in an Assessment Engineering framework.

As described in Chapter 2, an AE encourages the treatment of dimensionality to be largely addressed proactively in test development through the development of principled multidimensional information (PMI) that exhibits "simple structure". By specifying the number of traits of interest and building distinct sets of items to measure only a single specified trait, a set of separate unidimensional calibrations, one for each district trait will define the latent space for each trait, much as the same as the suggested scaling method in this study. If CPE are to constitute one or more of these distinct traits, any LID due to shared context or scoring association could be resolved by applying a polytomous scoring method.

Limitations and Future Studies

The current research is only a first step at systemically examining the sources and potential approaches to limit their influence. In general, simulation studies provide a convenient method to indentify issues that may be of interest and can only practically be examined under a finite set of conditions. In this study, a limited set of test format conditions were examined that represent some typical test situations where CPE and SR are both included. Future simulation studies may expand the test format variables to

include conditions that represent the most extreme situation possibly observed in practice (e.g. very small sample sizes, a single CPE per form, very short tests, etc.). Additionally, more complex testing environments, like computer adaptive test settings, could be simulated.

Other simulation studies might address the structure of the polytomous score units that are formed to address associations among tasks on the same CPE. In this study, all CPEs were designed to contain exactly the same number of MOs. It is likely that in operational tests, CPEs would be widely variable in the number of tasks they include. Another follow-up simulation study might attempt to create CPEs of different lengths, yielding different numbers of dichotomous MOs and/or polytomous score units. Expanding this idea would allow the polytomous score units that are created to contain differing number of score categories, unlike this study which constrained each polytomous unit formed to contain exactly four score categories.

As mentioned previously, the process used to create polytomous items was only effective in combating the affects of scoring factors, as items related to the same scoring factors were summed into a single item. The simulation in this study also linked all of the items within a CPE to a common context factor. Because the polytomous scoring method did not deal with this context factor, some LID due to these factors was likely observed among the polytomous items. A follow-up study could be designed that treated the context and scoring factors as separate conditions, and residual correlations compared across scoring methods that addresses one or both of these factors.

One reason this study was conducted using a simulation was to allow direct control over the amounts and types of influence various factors have on item responses. This approach, while useful for looking at particular aspects of a phenomena, will always present a limited view of the phenomena as it would appear in practice. In this study, the simulation allows for the creation and manipulation of two sources of construct irrelevant variance. Because these conditions were the only known sources of LID that could create association between two items within the same CPE, appropriate methods could be used to created polytomous items to address this source of LID. In reality, the situation is likely to be more complex, as there are a great number of factors that might influence responses. Further, the number and influence of these factors, coupled with the interaction among factors across various facets of the testing populations are likely complex and, of course, unknown. This makes the process of simply applying one of the methods employed in the study less clear, possibly leading to very different results. The logical follow-up study (or studies) is to apply the methods to outcomes from testing programs that contain a test format similar to that in the simulation. Such real data studies could demonstrate the effectiveness of the methods employed in this study. Other real data studies might examine the residual covariance structures at various places along the latent scale (e.g. at various cut points, or for groups of different abilities).

However, real data application and/or simulation studies must go beyond just describing the changes in residual covariance structures under different scaling and scoring procedures. Studies are needed that link the magnitude residual covariance and distribution of residual covariance to determine the real impact on testing outcomes—that

is, at what point does the residual covariance begin to affect the reported scores and/or item parameters in ways that compromise the fairness and validity of the scores. Q_3[12], though widely used a descriptive measure of LID, has no real guidelines for interpretation of magnitude of LID. An effect size study for this statistic is long overdue and would greatly assist in determining what the practical effects of LID.

This study used only three scaling methods. The methods selected were specifically chosen as they were practices that are operationally feasible. A logically extension of this research is to look at how the residual covariance manifests under other scaling models. These other models may include models for testlets, compensatory and non-compensatory MIRT models, or bifactor models. Of particular interest are methods that can be used when the number of items to be scaled is too small to be used in an IRT context (e.g. item level factor analysis or collateral information, as mentioned above).

[12]Under certain conditions in this study (e.g. Shorter tests that include both dichotomous and polytomous response), the expected value of Q_3 as presented in the literature did not match up with the empirical results. An empirical investigation that manipulates test format variables when no residual covariance is needed to define what the expectations before any attempt to evaluate

REFERENCES

Ackerman, T. A. (1987). *The Robustness of LOGIST and BILOG IRT Estimation Programs to Violations of Local Independence*. Paper presented at the Annual Meeting of the American Educational Research Association.

Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Prospect Heights, IL: Waveland Press, Inc.

Bachman, L. (2002). Alternative Interpretations of Alternative Assessments: Some Validity Issues in Educational Performance Assessments. *Educational Measurement: Issues & Practice, 21*(3), 5-18.

Bejar, I. I. (1988). An Approach to Assessing Unidimensionality Revisited. *Applied Psychological Measurement, 12*(4), 377.

Bennett, R. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice, 18*(3), 5-12.

Bennett, R. (2001). How the Internet Will Help Large-Scale Assessment Reinvent Itself. *Education Policy Analysis Archives, 9*(5).

Bennett, R., Bartram, D., & Hambleton, R. K. (2006). Inexorable and Inevitable: The Continuing Story of Technology and Assessment. In *Computer-based testing and*

the Internet: Issues and advances. (pp. 201-217). New York, NY, US: John Wiley & Sons Ltd.

Bennett, R., Goodman, M., Hessinger, J., Kahn, H., Ligget, J., Marshall, G., et al. (1999). Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior, 15*(3), 283-294.

Bennett, R., Jenkins, F., Persky, H., & Weiss, A. (2003). Assessing Complex Problem Solving Performances. *Assessment in Education: Principles, Policy & Practice, 10*(3), 347-359.

Bennett, R., Morley, M., & Quardt, D. (2000). Three response types for broadening the conception of mathematical problem solving in computerized tests. *Applied Psychological Measurement, 24*(4), 294-309.

Bennett, R., Steffen, M., Singley, M. K., & Morley, M. (1997). Evaluating an automatically scorable, open-ended response type for measuring mathematical reasoning in computer-adaptive tests. *Journal of Educational Measurement, 34*(2), 162-179.

Bennett, R., & Wadkins, J. R. (1995). Interactive performance assessment in computer science: The advanced placement computer science (APCS) practice system. *Journal of Educational Computing Research, 12*(4), 363-378.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. Lord & M. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*(2), 153-168.

Bridgeman, B., Bejar, I. I., & Friedman, D. (1999). Fairness issues in a computer-based architectural licensure examination. *Computers in Human Behavior, 15*(3), 419-440.

Burstein, J., Kukich, K., Wolff, S., Lu, C., & Chodorow, M. (2001). *Enriching Automated Essay Scoring Using Discourse Marking*.

Camara, W. (2003). *Scoring the Essay on the Writing Section*. New York: The College Board.

Campbell, D. T., & Fiske, D. W. (1959). Converegent and divergent validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Chen, W.-H., & Thissen, D. (1997). Local Dependence Indexes for Item Pairs Using Item Response Theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265.

Clauser, B. E. (2000). Recurrent issues and recent advances in scoring performance assessments. *Applied Psychological Measurement, 24*(4), 310-315.

Clauser, B. E., Kane, M. T., & Swanson, D. B. (2002). Validity issues for performance-based tests scored with computer-automated scoring systems. *Applied Measurement in Education, 15*(4), 413-432.

Clauser, B. E., Swanson, D. B., & Clyman, S. G. (1999). A comparison of the generalizability of scores produced by expert raters and automated scoring systems. *Applied Measurement in Education, 12*(3), 281-299.

Clyman, S. G., Melnick, D. E., Clauser, B. E., Mancall, E. L., & Bashook, P. G. (1995). Computer-based case simulations. In *Assessing clinical reasoning: The oral examination and alternative methods.* (pp. 139-149).

Coffman, W. E. (1971). Essay examinations. In R. L. Thorndike (Ed.), *Educational Measurement* (2 ed., pp. 271-302). Washington, D.C.: American Council on Education.

De Champlain, A. F. (1996). The Effect of Multidimensionality on ORT True-Score Equating for Subgroups of Examinees. *Journal of Educational Measurement, 33*(2), 181-201.

DeMars, C. E. (2006). Application of the Bi-Factor Multidimensional Item Response Theory Model to Testlet-Based Tests. *Journal of Educational Measurement, 43*(2), 145-168.

DeVore, R. (2004). *Consideration in the Development of Accounting Simulations*. Ewing, NJ: American Institute of Certified Public Accountants.

Drasgow, F., Luecht, R. M., & Bennett, R. (2006). Technology and Testing. In R. L. Brennan (Ed.), *Educational Measurement* (4 ed., pp. 471-515). Washington, DC: American Council on Education/Praeger Publishers.

Educational Testing Service, P. N. J. (2007). *TOEFL iBT 2007-2008 Bulliten* Princeton: Educational Testing Service.

Ferrara, S., & et al. (1997). Contextual Characteristics of Locally Dependent Open-Ended Item Clusters in a Large-Scale Performance Assessment. *Applied Measurement in Education, 10*(2), 123.

Ferrara, S., Huynh, H., & Michaels, H. (1999). Contextual explanations of local dependence in item clusters in a large scale hands-on science performance assessment. *Journal of Educational Measurement, 36*(2), 119-140.

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., et al. (2007). Full-Information Item Bifactor Analysis of Graded Response Data. *Applied Psychological Measurement, 31*(1), 4-19.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*(3), 423-436.

Goodman, J. T., & Luecht, R. M. (2007). *Assessing the Dependencies and Dimensionality of the Simulations on the Uniform Certified Public Accounting Examination* (No. W0703). Ewing: American Institute of Certified Public Accountants.

Goodman, J. T., Luecht, R. M., & Zhang, Y. (2008). *An Examination of the Magnitude of Residual Covariance for Complex Performance Assessments Under Various Scoring and Scaling Methods*. Paper presented at the National Council on Measurement in Education.

Habing, B., Finch, H., & Roberts, J. S. (2005). A Q-sub-3 Statistic for Unfolding Item Response Theory Models: Assessment of Unidimensionality With Two Factors and Simple Structure. *Applied Psychological Measurement, 29*(6), 457-471.

Haladyna, T. M. (1992). Context-dependent item sets. *Educational Measurement: Issues and Practice, 11*(1), 21-25.

Hambleton, R. K. (2000). Advances in Performance Assessment Methodolgy. *Applied Psychological Measurement, 24*(4), 291-293.

Hambleton, R. K., Jaeger, R. M., Plake, B. S., & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement, 24*(4), 355-366.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston, MA: Kluwer-Nijhoff Publishing.

Hanson, B. (2003). IRT Control Language (Version 1.02). Iowa City, IA.

Hardy, R. A. (1995). Examining the costs of performance assessment. *Applied Measurement in Education, 8*, 121-134.

Hartog, P., & Rhodes, E. C. (1936). *The marks of examiners*. London: MacMillian.

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (pp. 17-64). Washington, D.C.: American Council on Education/Praeger Publishers.

Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice, 18*(2), 5-17.

Lane, S. (2005). *Status and future directions for performance assessments in education*. Paper presented at the American Educational Research Association.

Lee, Y.-W. (2001). *The Essay Scoring and Scorer Reliability in TOEFL CBT*. Paper presented at the National Council on Measurement in Education (from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED455253&site=ehost-live

Lee, Y. W. (2004). Examining Passage-Related Local Item Dependence (LID) and Measurement Construct using Q3 Statistics in an EFL Reading Comprehension Test. *Language Testing, 21*(1), 74.

Li, Y., Bolt, D. M., & Fu, J. (2006). A Comparison of Alternative Models for Testlets. *Applied Psychological Measurement, 30*(1), 3-21.

Lin, Y.-H., Kim, S.-H., & Cohen, A. S. (2006). Local Dependence Indices and Detection Investigation for Polytmous Items Unpublished Paper.

Lindquest, E. F. (Ed.). (1951). *Educational Measurement* (1 ed.). Washington, DC: American Council on Education.

Linn, R. L., & Burton, E. (1994). Performance-based assessment: Implications of task specificity. *Educational Measurement: Issues & Practice, 13*(1).

Lord, F. (1952). A theory of test scores. *Psychometric Monographs*(7).

Lord, F. (1980). *Applications of Item Repsonse Theory to Practical Testing Problems*. Hillsdale, NJ: Lawerence Erlbaum Associates.

Lord, F., & Novick, M. (Eds.). (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley Publishing Company.

Luecht, R. M. (2001). *Capturing, Codifying and Scoring Complex Data for Innovative, Computer-Based Items.* Paper presented at the Annual Meeting of the National Council on Measurement in Education.

Luecht, R. M. (2004). MIRTGEN (Version 2). Greensboro, NC.

Luecht, R. M. (2005). Item Analysis. In B. Everitt & D. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science.* West Sussex, UK: John Wiley & Sons, Ltd.

Luecht, R. M. (2006). Engineering the Test: Principled Item Design to Automated Test Assembly. Invited special event presentation at the Annual Meeting of the Society for Industrial and Organizational Psychology.

Luecht, R. M. (2007a). Assessment Engineering: An Integrated Approach to Test Design, Development, Assembly, and Scoring., *Invited keynote and workshop presented at the Performance Testing Council Summit.* Scottsdale, AZ.

Luecht, R. M. (2007b). Emerging Topics (in CBT and CAT), *Invited keynote presentation at the 2007 GMAC® Conference on Computerized Adaptive Testing*. Minneapolis, MN.

Luecht, R. M., & Clauser, B. E. (2002). Test models for complex computer-based testing. In C. Mills, M. Potenza, J. Fremer & W. Ward (Eds.), *Computer-based Testing: Building the Foundation for Future Assessments* (pp. 67-88). Mahwah, NJ: Erlbaum.

Luecht, R. M., Gierl, M. J., & Huff, K. (2006). *Scalability and the Development of Useful Diagnostic Scores*. Paper presented at the Annual Meeting of the National Council on Measurement in Education.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Masters, G. N. (1988). The analysis of partial credit scoring. *Applied Measurement in Education, 1*, 279-297.

McDonald, R. P. (1981). The dimensionality of test and items. *British Journal of Mathematical and Statistical Psychology, 34*, 100-117.

McDonald, R. P. (1982a). Linear versus nonlinear models in latent trait theory. *Applied Psychological Measurement, 6*, 379-396.

McDonald, R. P. (1982b). *Unidimensional and Multidimensional Models for Item Response Theory.* Paper presented at the Item Response Theory and Computerized Adaptive Testing Conference, Wayzata, MN.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.

McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement, 24*(2), 99-114.

Messick , S. (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments. *Educational Researcher, 23*(2), 13.

Messick, S. (1995). Standards of Validity and the Validity of Standards in Performance Assessment. *Educational Measurement: Issues and Practice, 14*(4), 5.

Messick, S., & Hakel, M. D. (1998). Alternative modes of assessment, uniform standards of validity. In *Beyond multiple choice: Evaluating alternatives to traditional testing for selection.* (pp. 59-74). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Miller, M. D., & Linn, R. L. (2000). Validation of Performance-Based Assessment. *Applied Psychological Measurement, 24*(4), 367-378.

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement, 16*(2), 159.

Muraki, E., Hombo, C. M., & Lee, Y.-W. (2000). Equating and linking of performance assessments. *Applied Psychological Measurement, 24*(4), 325-337.

Nandakumar, R. (1991). Traditional dimensionality versus essential dimensionality. *Journal of Educational Measurement, 28*(2), 99-117.

Nandakumar, R. (1993). Assessing essential unidimensionality of real data. *Applied Psychological Measurement, 17*(1), 29-38.

Nandakumar, R., & Junker, B. W. (1993). *Estimation of Latent Ability Distributions under Essential Unidimensionality*.

Niemi, D., Baker, E. L., & Sylvester, R. M. (2007). Scaling Up, Scaling Down: Seven Years of Performance Assessment Development in the Nation's Second Largest School District. *Educational Assessment, 12*(3-4), 195.

Ostini, R., & Nering, M. (2006). *Polytomous Item Response Models* (Vol. 144). Thousands Oaks, NJ: Sage Publications.

Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. In W. J. Van der Linden & C. Glas (Eds.), *Computerized adaptive testing: Theory and practice*. Boston, MA: Kluwer.

Pearson, P. D., Calfee, R. C., Webb, P. L. W., & Fleischer, S. (2002). *The Role of Performance-Based Assessments in Large-Scale Accountability Systems: Lessons Learned from the Inside*. Washington, DC: Council of Chief State School Officers.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Oxford, England: Nielsen & Lydiche.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9*(4), 401-412.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*, 25-36.

Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement, 15*(4), 361-373.

Reese, L. M. (1995). *The Impact of Local Dependencies on Some LSAT Outcomes. Statistical Report. LSAC Research Report Series*.

Reese, L. M., & Pashley, P. J. (1999). *Impact of Local Item Dependence on True-Score Equating. LSAC Research Report Series*.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools

for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49*(3), 425-435.

Rudner, L., & Gagne, P. (2001). *An Overview of Three Approaches to Scoring Written Essays by Computer. ERIC Digest*.

Samejima, F. (1969). Estimation of latent ability using a repsonse pattern of graded responses. *Psychometrika, Monograph Supplements No. 18*.

Samejima, F. (1976). *Graded response model of the latent trait theory and tailored testing*. Paper presented at the First conference on computerized adaptive testing.

Sireci, S. G., & et al. (1991). On the Reliability of Testlet-Based Tests. *Journal of Educational Measurement, 28*(3), 237.

Sireci, S. G., & Rizavi, S. (2000). *Comparing Computerized and Human Scoring of Students' Essays*.

Stalnaker, J. M. (1951). The essay type of examination. In E. F. Lindquest (Ed.), *Educational Measurement* (1 ed., pp. 495-530). Washington, D.C.: American Council on Education.

Stark, S., Chernyshenko, O., & Drasgow, F. (2002). *Investigating the Effects of Local Dependence on the Accuracy of IRT Ability Estimation*. Ewing, NJ: American Institute of Certified Public Accountants.

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika, 55*(2), 293-325.

Stout, W. F., Habing, B., Douglas, J., & Kim, H. R. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement, 20*(4), 331-354.

Stout, W. F., Habing, B., Douglas, J., Kim, H. R., Rousses, L., & Zhang, J. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement, 20*, 331-354.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item repsonse models. *Psychometrika, 51*, 567-577.

Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement, 26*, 247-260.

Thompson, T. D., & Pommerich, M. (1996). *Examining the Sources and Effects of Local Dependence*. Paper presented at the American Educational Reserach Assocaition.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*, 529-554.

Thurstone, L. L. (1947). *Multiple Factor Analysis.* Chicago: University of Chicago Press.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Tucker, L. R. (1946). Maximum validity of a test with equivilent items. *Psychometrika, 11*, 1-13.

van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item repsonse theory*. New York: Springer-Verlag.

Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education, 6*(2), 103-118.

Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice, 15*(1), 22-29.

Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement, 37*(3), 203-220.

Wang, W.-C., Cheng, Y.-Y., & Wilson, M. (2005). Local Item Dependence for Items Across Tests Connected by Common Stimuli. *Educational and Psychological Measurement, 65*(1), 5-27.

Wiggins, G. (1993). Assessment: Authenticity, Context, and Validity. *Phi Delta Kappan, 75*(3ov), 200.

Wirth, R. J., & Edwards, M. C. (2007). Item Factor Analysis: Current Approaches and Future Directions. *Psychological Methods, 12*(1), 58-79.

Yan, J. W. (1997). *Examining Local Item Dependence Effects in a Large-Scale Science Assessment by a Rasch Partial Credit Model*.

Yao, L., & Schwarz, R. D. (2006). A Multidimensional Partial Credit Model With Associated Item and Test Statistics: An Application to Mixed-Format Tests. *Applied Psychological Measurement, 30*(6), 469-492.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125-145.

Yen, W. M. (1993). Scaling Performance Assessments: Strategies for Managing Local Item Dependence. *Journal of Educational Measurement, 30*(3), 187.

Yen, W. M. (1997). The technical quality of performance assessments: Standard errors of percents of pupils reaching standards. *Educational Measurement: Issues and Practice, 16*(3), 5-15.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2001). *Effects of Local Item Dependence on the Validity of IRT Item, Test, and Ability Statistics. MCAT Monograph*.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2002). Identification and evaluation of local item dependencies in the Medical College Admissions Test. *Journal of Educational Measurement, 39*(4), 291-309.

Zhang, J., & Stout, W. F. (1999). Conditional Covariance Structure of Generalized Compensatory Multidimensional Item. *Psychometrika, 64*(2), 129-152.

APPENDIX A: EXAMPLE COMMAND LANGUAGE FOR MIRTGEN

Example A1: Command language for study condition with 3000 test-takers, 120 items, 50% CPE items, high contextual dependencies, and a correlation 0.7 between traits

```
3000
24
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0
0.7 1.0
0.0 0.0 1.0
0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0
120
300012050hh.1.itm
300012050hh.1
```

Example A2: True item parameter files for data generation for above command file

```
1.10,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.52,0.17
0.96,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.45,0.08
0.90,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.39,0.11
0.85,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.64,0.14
1.00,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.40,0.12
1.12,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.51,0.03
0.99,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.51,0.15
0.90,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.14,0.17
0.98,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.24,0.16
1.11,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.05,0.12
0.92,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.44,0.01
0.92,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1.14,0.11
0.99,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.81,0.06
1.19,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.39,0.16
1.13,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.40,0.12
0.90,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.19,0.14
1.17,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.61,0.08
1.03,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.20,0.05
1.01,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.15,0.05
1.11,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.30,0.16
1.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1.06,0.12
1.00,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.56,0.03
1.15,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.19,0.17
1.01,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.11,0.19
1.04,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.22,0.12
1.12,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.39,0.01
1.02,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1.06,0
0.99,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.16,0.06
0.98,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1.67,0.16
1.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2.02,0.14
0.98,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.04,0.17
0.92,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.87,0.11
0.87,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1.28,0.19
0.89,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.74,0.18
1.30,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.17,0.07
```

```
1.00,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.40,0.09
1.07,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.62,0.02
1.14,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.09,0.14
1.00,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.87,0.15
0.93,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.69,0.10
1.08,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.07,0.13
0.81,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.32,0.18
0.91,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.26,0.19
0.91,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.61,0.15
1.00,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.09,0.20
1.01,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.12,0.04
0.89,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.49,0.05
1.01,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.30,0.10
1.00,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.45,0.08
1.04,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.21,0.19
1.10,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.91,0.01
0.85,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.69,0.06
0.93,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.34,0.16
1.07,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.16,0.19
0.95,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.68,0.09
1.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.63,0.01
1.11,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.70,0.05
1.11,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1.19,0.10
0.85,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.52,0.12
1.04,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.24,0.02
0,0.89,0.62,0,0.80,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.17,0
0,1.14,0.63,0,0.99,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.06,0
0,0.86,0.65,0,0.98,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.61,0
0,0.91,0.62,0,0,1.18,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.17,0
0,1.07,0.63,0,0,1.09,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.14,0
0,1.05,0.65,0,0,1.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.08,0
0,0.91,0.62,0,0,0,0.97,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.25,0
0,0.93,0.63,0,0,0,1.17,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1.02,0
0,1.06,0.65,0,0,0,0.98,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.59,0
0,0.96,0.62,0,0,0,0,1.04,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.46,0
0,0.90,0.63,0,0,0,0,1.18,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.30,0
0,1.17,0.65,0,0,0,0,0.89,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.12,0
0,1.02,0.62,0,0,0,0,0,0.88,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.54,0
0,0.97,0.63,0,0,0,0,0,0.86,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.64,0
0,0.99,0.65,0,0,0,0,0,1.01,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.20,0
0,0.93,0.62,0,0,0,0,0,0,1.12,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2.66,0
0,1.06,0.63,0,0,0,0,0,0,0,0.92,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.18,0
0,1.02,0.65,0,0,0,0,0,0,0.95,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.50,0
0,1.07,0.62,0,0,0,0,0,0,0,0.87,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.04,0
0,1.09,0.63,0,0,0,0,0,0,0,0,0.94,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.55,0
0,1.03,0.65,0,0,0,0,0,0,0,0.88,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.47,0
0,0.97,0.62,0,0,0,0,0,0,0,0,1.11,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.74,0
0,1.03,0.63,0,0,0,0,0,0,0,1.06,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.84,0
0,1.01,0.65,0,0,0,0,0,0,0,0,0.97,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-0.83,0
0,1.06,0.62,0,0,0,0,0,0,0,0,0,1.11,0,0,0,0,0,0,0,0,0,0,0,0,0,0.44,0
0,1.20,0.63,0,0,0,0,0,0,0,0,1.16,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.51,0
0,0.98,0.65,0,0,0,0,0,0,0,0,0,0.88,0,0,0,0,0,0,0,0,0,0,0,0,0,0.85,0
0,0.97,0.62,0,0,0,0,0,0,0,0,0,0,0,0.92,0,0,0,0,0,0,0,0,0,0,0,-1.58,0
0,1.07,0.63,0,0,0,0,0,0,0,0,0,0,0,0,0.83,0,0,0,0,0,0,0,0,0,0,-0.02,0
0,1.13,0.65,0,0,0,0,0,0,0,0,0,0,0,0,0,1.15,0,0,0,0,0,0,0,0,0,0.18,0
0,1.03,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.99,0,0,0,0,0,0,0,0,0,-0.57,0
0,1.01,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.98,0,0,0,0,0,0,0,0,0.34,0
0,1.08,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.99,0,0,0,0,0,0,0,0,0.75,0
0,0.84,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.01,0,0,0,0,0,0,0,0,0.63,0
0,0.98,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.10,0,0,0,0,0,0,0,0,1.31,0
0,1.10,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.94,0,0,0,0,0,0,0,0,0.82,0
0,0.98,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.04,0,0,0,0,0,0,0,0,0.35,0
0,1.05,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.13,0,0,0,0,0,0,0,0,0.81,0
0,0.87,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.06,0,0,0,0,0,0,0,0,-0.22,0
0,0.79,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.98,0,0,0,0,0,0,0,0,-0.65,0
0,1.01,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.13,0,0,0,0,0,0,0,0,-0.38,0
0,1.00,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.97,0,0,0,0,0,0,0,0,1.26,0
0,0.89,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.13,0,0,0,0,0,0,0,0,0.02,0
0,0.95,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.82,0,0,0,0,0,0,0,0,1.19,0
0,1.08,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.18,0,0,0,0,0,0,0,0,0.39,0
0,1.02,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.12,0,0,0,0,0,0,0,0,0.05,0
0,1.00,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.94,0,0,0,0,0,0,0,0,-0.28,0
0,0.94,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.88,0,0,0,0,0,0,0,0,-0.55,0
0,1.03,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.98,0,0,0,0,0,0,0,0,-0.65,0
0,1.03,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.87,0,0,0,0,0,0,0,0,0.70,0
0,0.91,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.13,0,0,0,0,0,0,0,0,0.08,0
0,1.04,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.05,0,0,0,0,0,0,0,0,0.16,0
0,0.80,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.04,0,0,0,0,0,0,0,0,0.46,0
0,1.10,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.92,0,0,0,1.58,0
0,1.03,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.92,0,-0.40,0
0,1.09,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.87,0,0.43,0
0,0.91,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.85,0,-0.27,0
0,1.05,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.10,-0.32,0
0,0.85,0,0.69,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.07,-0.49,0
0,1.14,0,0.64,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1.02,-0.08,0
```

APPENDIX B: ESTIMATED ITEM PARAMETERS

Table 10: Estimated and True Selected-Response *a* Parameters for 1000 Test- Takers and 60 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	1.11	0.17	1.00	0.10	0.10	1.11	0.17	1.00	0.10	0.11
			Polytomous	1.11	0.17	1.00	0.10	0.10	1.11	0.17	1.00	0.10	0.11
		Composite	Dichotomous	1.34	0.20	1.00	0.10	0.33	1.55	0.25	1.00	0.10	0.55
			Polytomous	1.37	0.21	1.00	0.10	0.37	1.75	0.25	1.00	0.10	0.75
		Separate	Dichotomous	1.11	0.17	1.00	0.10	0.10	1.11	0.17	1.00	0.10	0.11
			Polytomous	1.11	0.17	1.00	0.10	0.10	1.11	0.17	1.00	0.10	0.11
	0.5	One	Dichotomous	1.10	0.17	1.00	0.10	0.10	1.11	0.17	1.01	0.10	0.10
			Polytomous	1.10	0.17	1.00	0.10	0.10	1.11	0.17	1.01	0.10	0.10
		Composite	Dichotomous	1.23	0.20	1.00	0.10	0.23	1.39	0.25	1.01	0.10	0.39
			Polytomous	1.32	0.20	1.00	0.10	0.31	1.56	0.25	1.01	0.10	0.55
		Separate	Dichotomous	1.10	0.17	1.00	0.10	0.10	1.11	0.17	1.01	0.10	0.10
			Polytomous	1.10	0.17	1.00	0.10	0.10	1.11	0.17	1.01	0.10	0.10
	0.7	One	Dichotomous	1.12	0.18	1.01	0.10	0.11	1.13	0.18	1.01	0.10	0.12
			Polytomous	1.12	0.18	1.01	0.10	0.11	1.13	0.18	1.01	0.10	0.12
		Composite	Dichotomous	1.15	0.18	1.01	0.10	0.15	1.11	0.18	1.01	0.10	0.10
			Polytomous	1.29	0.20	1.01	0.10	0.28	1.42	0.22	1.01	0.10	0.41
		Separate	Dichotomous	1.12	0.18	1.01	0.10	0.11	1.13	0.18	1.01	0.10	0.12
			Polytomous	1.12	0.18	1.01	0.10	0.11	1.13	0.18	1.01	0.10	0.12
	1.0	One	Dichotomous	1.11	0.17	1.00	0.10	0.10	1.11	0.18	1.00	0.10	0.11
			Polytomous	1.11	0.17	1.00	0.10	0.10	1.11	0.18	1.00	0.10	0.11
		Composite	Dichotomous	1.12	0.18	1.00	0.10	0.11	1.12	0.18	1.00	0.10	0.11
			Polytomous	1.20	0.18	1.00	0.10	0.20	1.27	0.19	1.00	0.10	0.27
		Separate	Dichotomous	1.11	0.17	1.00	0.10	0.10	1.11	0.18	1.00	0.10	0.11
			Polytomous	1.11	0.17	1.00	0.10	0.10	1.11	0.18	1.00	0.10	0.11
Low	0.2	One	Dichotomous	1.10	0.18	1.00	0.10	0.10	1.11	0.18	1.01	0.10	0.11
			Polytomous	1.10	0.18	1.00	0.10	0.10	1.11	0.18	1.01	0.10	0.11
		Composite	Dichotomous	1.53	0.24	1.00	0.10	0.53	0.84	0.19	1.01	0.10	-0.17
			Polytomous	1.38	0.22	1.00	0.10	0.38	2.08	0.27	1.01	0.10	1.07
		Separate	Dichotomous	1.10	0.18	1.00	0.10	0.10	1.11	0.18	1.01	0.10	0.11
			Polytomous	1.10	0.18	1.00	0.10	0.10	1.11	0.18	1.01	0.10	0.11
	0.5	One	Dichotomous	1.11	0.17	1.01	0.10	0.10	1.11	0.18	1.00	0.10	0.11
			Polytomous	1.11	0.17	1.01	0.10	0.10	1.11	0.18	1.00	0.10	0.11
		Composite	Dichotomous	1.20	0.20	1.01	0.10	0.20	0.94	0.19	1.00	0.10	-0.07
			Polytomous	1.33	0.21	1.01	0.10	0.32	1.52	0.26	1.00	0.10	0.52
		Separate	Dichotomous	1.11	0.17	1.01	0.10	0.10	1.11	0.18	1.00	0.10	0.11
			Polytomous	1.11	0.17	1.01	0.10	0.10	1.11	0.18	1.00	0.10	0.11
	0.7	One	Dichotomous	1.11	0.18	1.01	0.10	0.10	1.12	0.17	1.01	0.10	0.11
			Polytomous	1.11	0.18	1.01	0.10	0.10	1.12	0.17	1.01	0.10	0.11
		Composite	Dichotomous	1.12	0.17	1.01	0.10	0.11	0.99	0.16	1.01	0.10	-0.02
			Polytomous	1.25	0.19	1.01	0.10	0.24	1.26	0.19	1.01	0.10	0.25
		Separate	Dichotomous	1.11	0.18	1.01	0.10	0.10	1.12	0.17	1.01	0.10	0.11
			Polytomous	1.11	0.18	1.01	0.10	0.10	1.12	0.17	1.01	0.10	0.11
	1.0	One	Dichotomous	1.11	0.18	1.01	0.10	0.10	1.11	0.17	1.01	0.10	0.11
			Polytomous	1.11	0.18	1.01	0.10	0.10	1.11	0.17	1.01	0.10	0.11
		Composite	Dichotomous	1.11	0.18	1.01	0.10	0.11	1.12	0.18	1.01	0.10	0.11
			Polytomous	1.15	0.18	1.01	0.10	0.14	1.18	0.17	1.01	0.10	0.17
		Separate	Dichotomous	1.11	0.18	1.01	0.10	0.10	1.11	0.17	1.01	0.10	0.11
			Polytomous	1.11	0.18	1.01	0.10	0.10	1.11	0.17	1.01	0.10	0.11
None	0.2	One	Dichotomous	1.11	0.17	1.00	0.10	0.11	1.12	0.18	1.00	0.10	0.11
			Polytomous	1.11	0.17	1.00	0.10	0.11	1.12	0.18	1.00	0.10	0.11
		Composite	Dichotomous	1.52	0.22	1.00	0.10	0.52	0.50	0.12	1.00	0.10	-0.50
			Polytomous	1.43	0.21	1.00	0.10	0.42	2.22	0.29	1.00	0.10	1.22
		Separate	Dichotomous	1.11	0.17	1.00	0.10	0.11	1.12	0.18	1.00	0.10	0.11
			Polytomous	1.11	0.17	1.00	0.10	0.11	1.12	0.18	1.00	0.10	0.11
	0.5	One	Dichotomous	1.11	0.18	1.00	0.10	0.11	1.12	0.18	1.01	0.10	0.11
			Polytomous	1.11	0.18	1.00	0.10	0.11	1.12	0.18	1.01	0.10	0.11

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
		Composite	Dichotomous	1.19	0.19	1.00	0.10	0.19	0.81	0.16	1.01	0.10	-0.19
			Polytomous	1.32	0.21	1.00	0.10	0.32	1.38	0.26	1.01	0.10	0.37
		Separate	Dichotomous	1.11	0.18	1.00	0.10	0.11	1.12	0.18	1.01	0.10	0.11
			Polytomous	1.11	0.18	1.00	0.10	0.11	1.12	0.18	1.01	0.10	0.11
	0.7	One	Dichotomous	1.11	0.18	1.01	0.10	0.11	1.12	0.18	1.01	0.10	0.11
			Polytomous	1.11	0.18	1.01	0.10	0.11	1.12	0.18	1.01	0.10	0.11
		Composite	Dichotomous	1.10	0.17	1.01	0.10	0.09	0.94	0.15	1.01	0.10	-0.07
			Polytomous	1.22	0.19	1.01	0.10	0.21	1.12	0.17	1.01	0.10	0.11
		Separate	Dichotomous	1.11	0.18	1.01	0.10	0.11	1.12	0.18	1.01	0.10	0.11
			Polytomous	1.11	0.18	1.01	0.10	0.11	1.12	0.18	1.01	0.10	0.11
	1.0	One	Dichotomous	1.10	0.17	1.01	0.10	0.10	1.11	0.18	1.00	0.10	0.11
			Polytomous	1.10	0.17	1.01	0.10	0.10	1.11	0.18	1.00	0.10	0.11
		Composite	Dichotomous	1.11	0.18	1.01	0.10	0.10	1.13	0.18	1.00	0.10	0.12
			Polytomous	1.12	0.17	1.01	0.10	0.11	1.14	0.16	1.00	0.10	0.13
		Separate	Dichotomous	1.10	0.17	1.01	0.10	0.10	1.11	0.18	1.00	0.10	0.11
			Polytomous	1.10	0.17	1.01	0.10	0.10	1.11	0.18	1.00	0.10	0.11

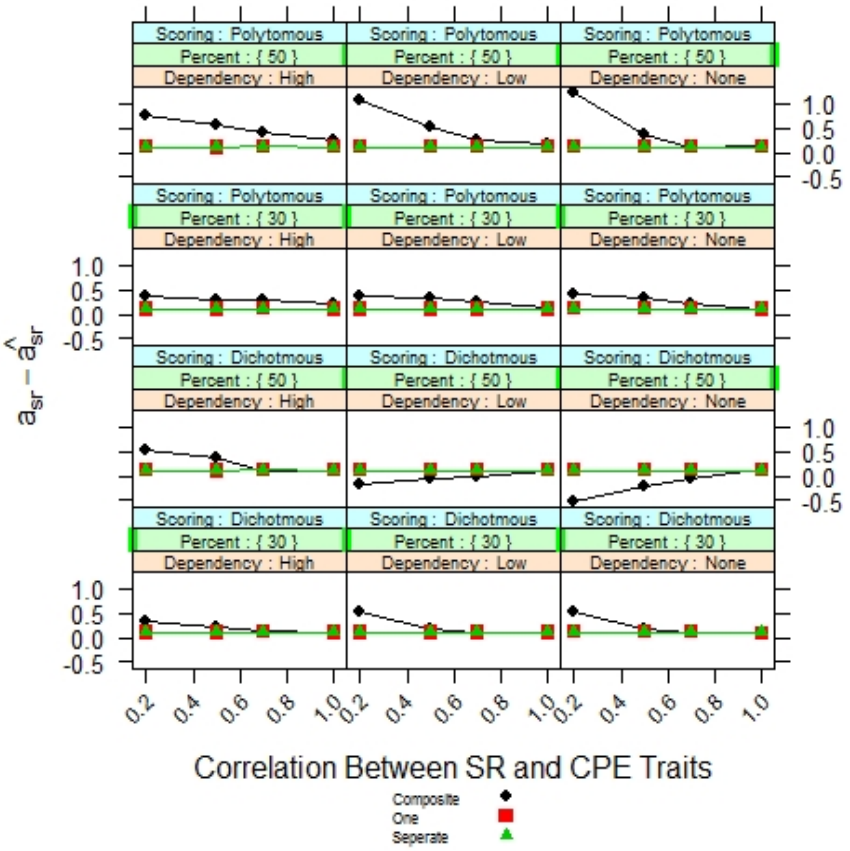

Figure 10: Difference Between Estimated and True Selected Response "a" Parameters for Sample Size of 1000 and 60 Items

Table 11: Average Estimated and True Selected Response *a* Parameter for 1000 Test-Takers and 120 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	1.10	0.17	1.01	0.10	0.09	1.10	0.18	1.00	0.10	0.10
			Polytomous	1.10	0.17	1.01	0.10	0.09	1.10	0.18	1.00	0.10	0.10
		Composite	Dichotomous	1.40	0.20	1.01	0.10	0.39	1.95	0.30	1.00	0.10	0.95
			Polytomous	1.41	0.21	1.01	0.10	0.40	1.86	0.26	1.00	0.10	0.85
		Separate	Dichotomous	1.10	0.17	1.01	0.10	0.09	1.10	0.18	1.00	0.10	0.10
			Polytomous	1.10	0.17	1.01	0.10	0.09	1.10	0.18	1.00	0.10	0.10
	0.5	One	Dichotomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.00	0.10	0.09
			Polytomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.00	0.10	0.09
		Composite	Dichotomous	1.20	0.18	1.00	0.10	0.19	1.29	0.22	1.00	0.10	0.28
			Polytomous	1.34	0.20	1.00	0.10	0.33	1.56	0.24	1.00	0.10	0.55
		Separate	Dichotomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.00	0.10	0.09
			Polytomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.00	0.10	0.09
	0.7	One	Dichotomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.09
		Composite	Dichotomous	1.12	0.17	1.01	0.10	0.11	1.07	0.17	1.01	0.10	0.07
			Polytomous	1.28	0.19	1.01	0.10	0.28	1.41	0.21	1.01	0.10	0.41
		Separate	Dichotomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.09
	1.0	One	Dichotomous	1.10	0.16	1.00	0.10	0.09	1.11	0.18	1.00	0.10	0.10
			Polytomous	1.10	0.16	1.00	0.10	0.09	1.11	0.18	1.00	0.10	0.10
		Composite	Dichotomous	1.11	0.17	1.00	0.10	0.10	1.11	0.18	1.00	0.10	0.10
			Polytomous	1.20	0.17	1.00	0.10	0.20	1.28	0.19	1.00	0.10	0.28
		Separate	Dichotomous	1.10	0.16	1.00	0.10	0.09	1.11	0.18	1.00	0.10	0.10
			Polytomous	1.10	0.16	1.00	0.10	0.09	1.11	0.18	1.00	0.10	0.10
Low	0.2	One	Dichotomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.09
		Composite	Dichotomous	1.52	0.22	1.01	0.10	0.51	0.52	0.11	1.01	0.10	-0.48
			Polytomous	1.45	0.21	1.01	0.10	0.44	2.12	0.28	1.01	0.10	1.11
		Separate	Dichotomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.09
	0.5	One	Dichotomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.00	0.10	0.10
			Polytomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.00	0.10	0.10
		Composite	Dichotomous	1.20	0.19	1.00	0.10	0.20	1.07	0.21	1.00	0.10	0.06
			Polytomous	1.33	0.20	1.00	0.10	0.33	1.58	0.27	1.00	0.10	0.57
		Separate	Dichotomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.00	0.10	0.10
			Polytomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.00	0.10	0.10
	0.7	One	Dichotomous	1.10	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.10	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
		Composite	Dichotomous	1.09	0.16	1.00	0.10	0.08	0.97	0.15	1.01	0.10	-0.03
			Polytomous	1.24	0.18	1.00	0.10	0.24	1.25	0.18	1.01	0.10	0.25
		Separate	Dichotomous	1.10	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.10	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
	1.0	One	Dichotomous	1.09	0.16	1.01	0.10	0.08	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.09	0.16	1.01	0.10	0.08	1.10	0.17	1.01	0.10	0.09
		Composite	Dichotomous	1.09	0.17	1.01	0.10	0.08	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.14	0.16	1.01	0.10	0.13	1.18	0.17	1.01	0.10	0.17
		Separate	Dichotomous	1.09	0.16	1.01	0.10	0.08	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.09	0.16	1.01	0.10	0.08	1.10	0.17	1.01	0.10	0.09
None	0.2	One	Dichotomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.00	0.10	0.09
			Polytomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.00	0.10	0.09
		Composite	Dichotomous	1.64	0.25	1.01	0.10	0.63	0.54	0.16	1.00	0.10	-0.46
			Polytomous	1.51	0.22	1.01	0.10	0.50	2.24	0.30	1.00	0.10	1.24
		Separate	Dichotomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.00	0.10	0.09
			Polytomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.00	0.10	0.09
	0.5	One	Dichotomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.10
			Polytomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.10
		Composite	Dichotomous	1.26	0.20	1.01	0.10	0.25	0.74	0.14	1.01	0.10	-0.26
			Polytomous	1.36	0.20	1.01	0.10	0.36	1.28	0.21	1.01	0.10	0.27
		Separate	Dichotomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.10
			Polytomous	1.09	0.17	1.01	0.10	0.09	1.10	0.17	1.01	0.10	0.10
	0.7	One	Dichotomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
		Composite	Dichotomous	1.07	0.17	1.00	0.10	0.07	0.93	0.15	1.01	0.10	-0.08
			Polytomous	1.21	0.18	1.00	0.10	0.21	1.13	0.17	1.01	0.10	0.12

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
		Separate	Dichotomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
	1.0	One	Dichotomous	1.08	0.17	1.00	0.10	0.08	1.10	0.16	1.01	0.10	0.09
			Polytomous	1.08	0.17	1.00	0.10	0.08	1.10	0.16	1.01	0.10	0.09
		Composite	Dichotomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.11	0.16	1.00	0.10	0.10	1.14	0.15	1.01	0.10	0.14
		Separate	Dichotomous	1.08	0.17	1.00	0.10	0.08	1.10	0.16	1.01	0.10	0.09
			Polytomous	1.08	0.17	1.00	0.10	0.08	1.10	0.16	1.01	0.10	0.09

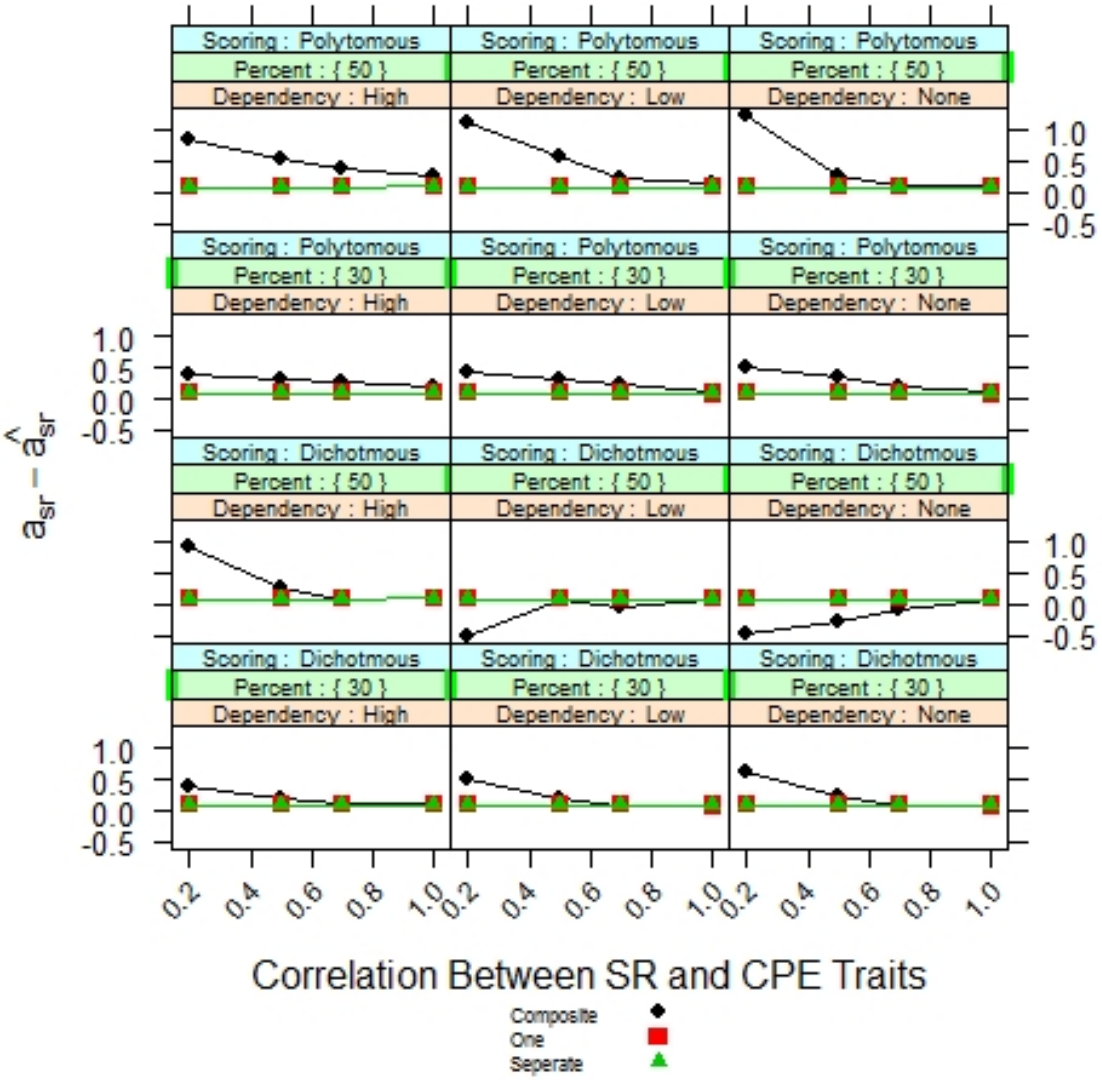

Figure 11: Difference Between Estimated and True Selected Response "a" Parameters for Sample Size of 1000 and 120 Items

Table 12: Average Estimated and True Selected Response *a* Parameter for 3000 Test-Takers and 60 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.08	0.14	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.08	0.14	1.01	0.10	0.07
		Composite	Dichotomous	1.74	0.20	1.00	0.10	0.74	1.92	0.20	1.01	0.10	0.91
			Polytomous	1.37	0.17	1.00	0.10	0.37	2.10	0.23	1.01	0.10	1.09
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.08	0.14	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.08	0.14	1.01	0.10	0.07
	0.5	One	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.01	0.10	0.07
		Composite	Dichotomous	1.30	0.17	1.00	0.10	0.30	2.07	0.31	1.01	0.10	1.07
			Polytomous	1.33	0.16	1.00	0.10	0.32	1.79	0.22	1.01	0.10	0.79
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.01	0.10	0.07
	0.7	One	Dichotomous	1.08	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
			Polytomous	1.08	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
		Composite	Dichotomous	1.15	0.14	1.00	0.10	0.15	1.10	0.13	1.00	0.10	0.10
			Polytomous	1.26	0.15	1.00	0.10	0.26	1.41	0.17	1.00	0.10	0.40
		Separate	Dichotomous	1.08	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
			Polytomous	1.08	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
	1.0	One	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.01	0.10	0.06
		Composite	Dichotomous	1.09	0.13	1.00	0.10	0.09	1.09	0.13	1.01	0.10	0.09
			Polytomous	1.16	0.14	1.00	0.10	0.16	1.23	0.15	1.01	0.10	0.23
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.01	0.10	0.06
Low	0.2	One	Dichotomous	1.07	0.12	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.07
			Polytomous	1.07	0.12	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.07
		Composite	Dichotomous	1.96	0.22	1.00	0.10	0.96	1.00	0.23	1.00	0.10	0.00
			Polytomous	1.41	0.16	1.00	0.10	0.41	2.47	0.25	1.00	0.10	1.46
		Separate	Dichotomous	1.07	0.12	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.07
			Polytomous	1.07	0.12	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.07
	0.5	One	Dichotomous	1.07	0.13	1.01	0.10	0.07	1.07	0.13	1.00	0.10	0.07
			Polytomous	1.07	0.13	1.01	0.10	0.07	1.07	0.13	1.00	0.10	0.07
		Composite	Dichotomous	1.34	0.18	1.01	0.10	0.33	1.38	0.24	1.00	0.10	0.38
			Polytomous	1.35	0.17	1.01	0.10	0.34	1.91	0.29	1.00	0.10	0.91
		Separate	Dichotomous	1.07	0.13	1.01	0.10	0.07	1.07	0.13	1.00	0.10	0.07
			Polytomous	1.07	0.13	1.01	0.10	0.07	1.07	0.13	1.00	0.10	0.07
	0.7	One	Dichotomous	1.06	0.13	1.00	0.10	0.07	1.08	0.13	1.01	0.10	0.07
			Polytomous	1.06	0.13	1.00	0.10	0.07	1.08	0.13	1.01	0.10	0.07
		Composite	Dichotomous	1.10	0.13	1.00	0.10	0.10	0.99	0.11	1.01	0.10	-0.02
			Polytomous	1.22	0.15	1.00	0.10	0.22	1.26	0.15	1.01	0.10	0.25
		Separate	Dichotomous	1.06	0.13	1.00	0.10	0.07	1.08	0.13	1.01	0.10	0.07
			Polytomous	1.06	0.13	1.00	0.10	0.07	1.08	0.13	1.01	0.10	0.07
	1.0	One	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.06	0.13	1.00	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.06	0.13	1.00	0.10	0.07
		Composite	Dichotomous	1.09	0.13	1.00	0.10	0.08	1.09	0.13	1.00	0.10	0.09
			Polytomous	1.12	0.13	1.00	0.10	0.12	1.15	0.13	1.00	0.10	0.15
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.06	0.13	1.00	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.06	0.13	1.00	0.10	0.07
None	0.2	One	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.08	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.08	0.13	1.01	0.10	0.07
		Composite	Dichotomous	1.89	0.22	1.00	0.10	0.88	0.53	0.14	1.01	0.10	-0.48
			Polytomous	1.55	0.18	1.00	0.10	0.54	2.57	0.25	1.01	0.10	1.56
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.08	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.08	0.13	1.01	0.10	0.07
	0.5	One	Dichotomous	1.07	0.13	1.01	0.10	0.06	1.08	0.13	1.00	0.10	0.07
			Polytomous	1.07	0.13	1.01	0.10	0.06	1.08	0.13	1.00	0.10	0.07
		Composite	Dichotomous	1.27	0.16	1.01	0.10	0.26	0.88	0.13	1.00	0.10	-0.12
			Polytomous	1.35	0.16	1.01	0.10	0.34	1.45	0.22	1.00	0.10	0.45
		Separate	Dichotomous	1.07	0.13	1.01	0.10	0.06	1.08	0.13	1.00	0.10	0.07
			Polytomous	1.07	0.13	1.01	0.10	0.06	1.08	0.13	1.00	0.10	0.07
	0.7	One	Dichotomous	1.08	0.13	1.01	0.10	0.07	1.07	0.13	1.00	0.10	0.07

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
		Composite	Polytomous	1.08	0.13	1.01	0.10	0.07	1.07	0.13	1.00	0.10	0.07
			Dichotomous	1.10	0.13	1.01	0.10	0.09	0.92	0.10	1.00	0.10	-0.08
		Separate	Polytomous	1.21	0.15	1.01	0.10	0.20	1.09	0.12	1.00	0.10	0.08
			Dichotomous	1.08	0.13	1.01	0.10	0.07	1.07	0.13	1.00	0.10	0.07
			Polytomous	1.08	0.13	1.01	0.10	0.07	1.07	0.13	1.00	0.10	0.07
	1.0	One	Dichotomous	1.08	0.17	1.00	0.10	0.08	1.10	0.16	1.01	0.10	0.09
			Polytomous	1.08	0.17	1.00	0.10	0.08	1.10	0.16	1.01	0.10	0.09
		Composite	Dichotomous	1.09	0.17	1.00	0.10	0.09	1.10	0.17	1.01	0.10	0.09
			Polytomous	1.11	0.16	1.00	0.10	0.10	1.14	0.15	1.01	0.10	0.14
		Separate	Dichotomous	1.08	0.17	1.00	0.10	0.08	1.10	0.16	1.01	0.10	0.09
			Polytomous	1.08	0.17	1.00	0.10	0.08	1.10	0.16	1.01	0.10	0.09

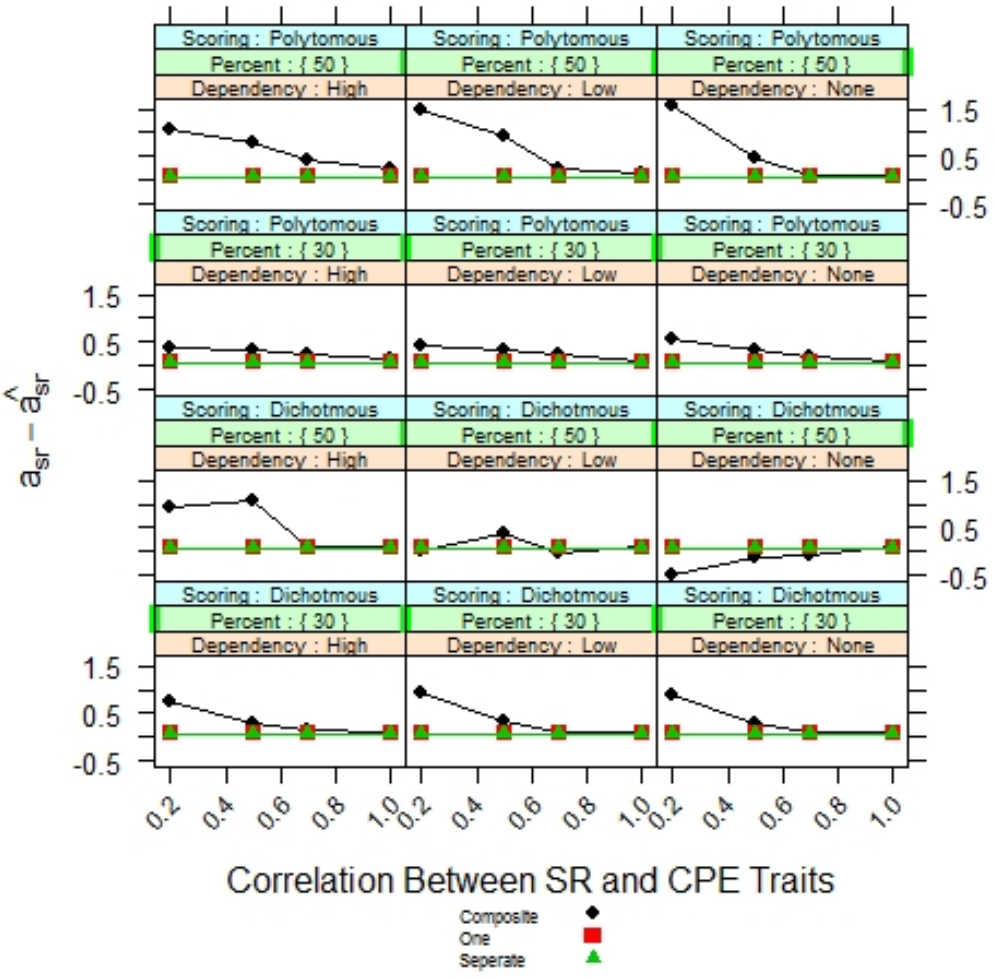

Figure 12: Difference Between Estimated and True Selected Response "a" Parameters for Sample Size of 3000 and 60 Items

Table 13: Average Estimated and True Selected Response *a* Parameter for 3000 Test-Takers and 120 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
		Composite	Dichotomous	1.84	0.20	1.00	0.10	0.84	2.50	0.25	1.01	0.10	1.50
			Polytomous	1.43	0.17	1.00	0.10	0.43	2.21	0.23	1.01	0.10	1.20
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
	0.5	One	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
		Composite	Dichotomous	1.28	0.15	1.00	0.10	0.28	1.90	0.34	1.00	0.10	0.90
			Polytomous	1.37	0.16	1.00	0.10	0.36	1.75	0.22	1.00	0.10	0.74
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
	0.7	One	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
		Composite	Dichotomous	1.12	0.13	1.00	0.10	0.12	1.06	0.12	1.00	0.10	0.05
			Polytomous	1.27	0.15	1.00	0.10	0.27	1.42	0.17	1.00	0.10	0.42
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.07	1.07	0.13	1.00	0.10	0.06
	1.0	One	Dichotomous	1.06	0.13	1.00	0.10	0.06	1.07	0.13	1.00	0.10	0.07
			Polytomous	1.06	0.13	1.00	0.10	0.06	1.07	0.13	1.00	0.10	0.07
		Composite	Dichotomous	1.08	0.13	1.00	0.10	0.07	1.08	0.12	1.00	0.10	0.07
			Polytomous	1.17	0.14	1.00	0.10	0.17	1.25	0.15	1.00	0.10	0.24
		Separate	Dichotomous	1.06	0.13	1.00	0.10	0.06	1.07	0.13	1.00	0.10	0.07
			Polytomous	1.06	0.13	1.00	0.10	0.06	1.07	0.13	1.00	0.10	0.07
Low	0.2	One	Dichotomous	1.07	0.12	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.12	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.06
		Composite	Dichotomous	2.02	0.22	1.00	0.10	1.01	1.17	0.26	1.01	0.10	0.16
			Polytomous	1.53	0.17	1.00	0.10	0.52	2.50	0.24	1.01	0.10	1.49
		Separate	Dichotomous	1.07	0.12	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.12	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.06
	0.5	One	Dichotomous	1.07	0.13	1.01	0.10	0.06	1.08	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.01	0.10	0.06	1.08	0.13	1.01	0.10	0.07
		Composite	Dichotomous	1.30	0.17	1.01	0.10	0.29	1.19	0.20	1.01	0.10	0.18
			Polytomous	1.36	0.16	1.01	0.10	0.36	1.81	0.28	1.01	0.10	0.81
		Separate	Dichotomous	1.07	0.13	1.01	0.10	0.06	1.08	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.01	0.10	0.06	1.08	0.13	1.01	0.10	0.07
	0.7	One	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
		Composite	Dichotomous	1.09	0.13	1.00	0.10	0.08	0.94	0.11	1.01	0.10	-0.06
			Polytomous	1.24	0.15	1.00	0.10	0.23	1.24	0.15	1.01	0.10	0.23
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
	1.0	One	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.06
		Composite	Dichotomous	1.07	0.13	1.00	0.10	0.07	1.08	0.13	1.01	0.10	0.07
			Polytomous	1.12	0.13	1.00	0.10	0.11	1.16	0.14	1.01	0.10	0.15
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.06
None	0.2	One	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
		Composite	Dichotomous	2.22	0.25	1.00	0.10	1.21	0.60	0.15	1.01	0.10	-0.40
			Polytomous	1.73	0.19	1.00	0.10	0.73	2.62	0.31	1.01	0.10	1.61
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.07	0.13	1.01	0.10	0.07
	0.5	One	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.06	0.13	1.00	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.06	0.13	1.00	0.10	0.06
		Composite	Dichotomous	1.28	0.16	1.00	0.10	0.28	0.78	0.12	1.00	0.10	-0.22
			Polytomous	1.38	0.16	1.00	0.10	0.37	1.42	0.23	1.00	0.10	0.42
		Separate	Dichotomous	1.07	0.13	1.00	0.10	0.06	1.06	0.13	1.00	0.10	0.06
			Polytomous	1.07	0.13	1.00	0.10	0.06	1.06	0.13	1.00	0.10	0.06
	0.7	One	Dichotomous	1.07	0.13	1.01	0.10	0.06	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.13	1.01	0.10	0.06	1.07	0.13	1.01	0.10	0.06
		Composite	Dichotomous	1.07	0.12	1.01	0.10	0.06	0.91	0.10	1.01	0.10	-0.10
			Polytomous	1.20	0.14	1.01	0.10	0.19	1.10	0.12	1.01	0.10	0.09
		Separate	Dichotomous	1.07	0.13	1.01	0.10	0.06	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.13	1.01	0.10	0.06	1.07	0.13	1.01	0.10	0.06

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
	1.0	One	Dichotomous	1.07	0.13	1.01	0.10	0.06	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.13	1.01	0.10	0.06	1.07	0.13	1.01	0.10	0.06
		Composite	Dichotomous	1.08	0.13	1.01	0.10	0.07	1.08	0.13	1.01	0.10	0.08
			Polytomous	1.09	0.12	1.01	0.10	0.08	1.11	0.13	1.01	0.10	0.11
		Separate	Dichotomous	1.07	0.13	1.01	0.10	0.06	1.07	0.13	1.01	0.10	0.06
			Polytomous	1.07	0.13	1.01	0.10	0.06	1.07	0.13	1.01	0.10	0.06

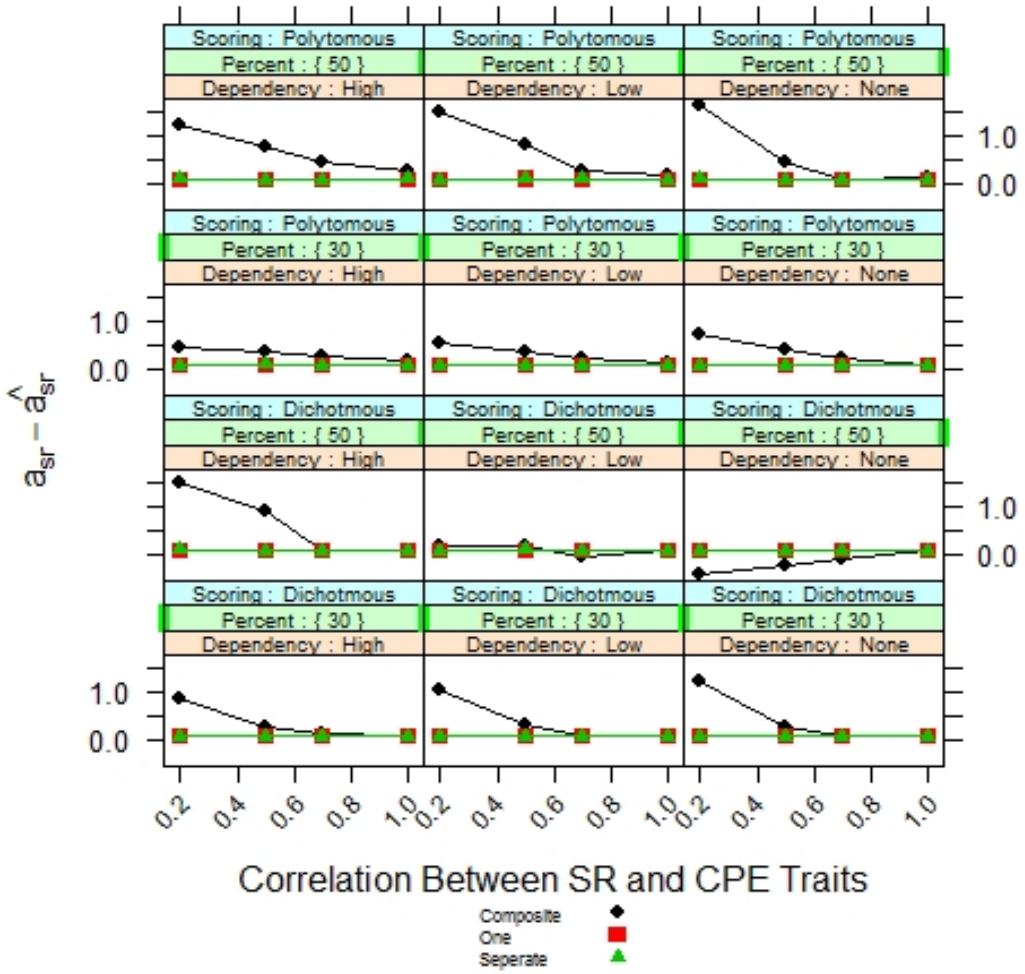

Figure 13: Difference Between Estimated and True Selected Response "a" Parameters for Sample Size of 3000 and 120 Items

Table 14: Average Estimated and True CPE *a* Parameter for 1000 Test- Takers and 60 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.62	0.10	1.01	0.15	-0.39	0.79	0.12	1.01	0.15	-0.21
			Polytomous	0.42	0.18	1.01	0.15	-0.59	0.51	0.18	1.01	0.15	-0.50
		Composite	Dichotomous	0.61	0.11	1.01	0.15	-0.39	0.78	0.13	1.01	0.15	-0.23
			Polytomous	0.47	0.20	1.01	0.15	-0.53	0.61	0.22	1.01	0.15	-0.40
		Separate	Dichotomous	1.14	0.37	1.01	0.15	0.13	1.02	0.22	1.01	0.15	0.02
			Polytomous	0.96	0.14	1.01	0.15	-0.05	1.01	0.14	1.01	0.15	0.00
	0.5	One	Dichotomous	0.79	0.12	1.02	0.15	-0.23	1.02	0.16	1.02	0.15	0.00
			Polytomous	0.67	0.18	1.02	0.15	-0.34	0.93	0.17	1.02	0.15	-0.09
		Composite	Dichotomous	0.78	0.12	1.02	0.15	-0.24	0.99	0.17	1.02	0.15	-0.03
			Polytomous	0.74	0.19	1.02	0.15	-0.27	1.08	0.19	1.02	0.15	0.06
		Separate	Dichotomous	1.12	0.21	1.02	0.15	0.11	1.11	0.20	1.02	0.15	0.09
			Polytomous	1.40	0.18	1.02	0.15	0.38	1.46	0.18	1.02	0.15	0.44
	0.7	One	Dichotomous	0.89	0.12	1.02	0.15	-0.13	1.16	0.17	1.01	0.14	0.15
			Polytomous	0.93	0.18	1.02	0.15	-0.10	1.39	0.18	1.01	0.14	0.38
		Composite	Dichotomous	0.89	0.13	1.02	0.15	-0.14	1.11	0.17	1.01	0.14	0.09
			Polytomous	1.01	0.19	1.02	0.15	-0.01	1.54	0.17	1.01	0.14	0.53
		Separate	Dichotomous	1.19	0.22	1.02	0.15	0.16	1.18	0.20	1.01	0.14	0.17
			Polytomous	1.84	0.18	1.02	0.15	0.82	1.86	0.18	1.01	0.14	0.85
	1.0	One	Dichotomous	0.31	0.18	1.01	0.15	-0.70	0.76	0.26	1.01	0.15	-0.25
			Polytomous	0.12	0.17	1.01	0.15	-0.89	0.15	0.17	1.01	0.15	-0.86
		Composite	Dichotomous	0.29	0.13	1.01	0.15	-0.72	1.05	0.20	1.01	0.15	0.04
			Polytomous	0.14	0.21	1.01	0.15	-0.87	0.20	0.25	1.01	0.15	-0.81
		Separate	Dichotomous	1.17	0.42	1.01	0.15	0.16	1.02	0.21	1.01	0.15	0.01
			Polytomous	0.96	0.13	1.01	0.15	-0.06	1.00	0.13	1.01	0.15	-0.01
Low	0.2	One	Dichotomous	0.36	0.14	1.01	0.14	-0.66	1.16	0.24	1.01	0.15	0.15
			Polytomous	0.18	0.18	1.01	0.14	-0.84	0.26	0.18	1.01	0.15	-0.76
		Composite	Dichotomous	0.42	0.12	1.01	0.14	-0.60	1.81	0.24	1.01	0.15	0.80
			Polytomous	0.21	0.22	1.01	0.14	-0.81	0.32	0.27	1.01	0.15	-0.69
		Separate	Dichotomous	1.11	0.20	1.01	0.14	0.10	1.11	0.19	1.01	0.15	0.10
			Polytomous	1.38	0.17	1.01	0.14	0.37	1.46	0.17	1.01	0.15	0.45
	0.5	One	Dichotomous	0.48	0.23	1.01	0.15	-0.53	1.54	0.23	1.02	0.16	0.53
			Polytomous	0.21	0.17	1.01	0.15	-0.80	0.39	0.18	1.02	0.16	-0.63
		Composite	Dichotomous	0.51	0.20	1.01	0.15	-0.50	2.06	0.29	1.02	0.16	1.05
			Polytomous	0.25	0.21	1.01	0.15	-0.76	0.48	0.29	1.02	0.16	-0.54
		Separate	Dichotomous	1.17	0.21	1.01	0.15	0.16	1.18	0.21	1.02	0.16	0.17
			Polytomous	1.83	0.17	1.01	0.15	0.82	1.88	0.19	1.02	0.16	0.86
	0.7	One	Dichotomous	0.49	0.08	1.01	0.15	-0.52	0.72	0.11	1.02	0.15	-0.30
			Polytomous	0.29	0.17	1.01	0.15	-0.72	0.36	0.17	1.02	0.15	-0.65
		Composite	Dichotomous	0.45	0.08	1.01	0.15	-0.56	0.79	0.14	1.02	0.15	-0.22
			Polytomous	0.33	0.20	1.01	0.15	-0.68	0.39	0.25	1.02	0.15	-0.63
		Separate	Dichotomous	1.15	0.38	1.01	0.15	0.14	1.04	0.22	1.02	0.15	0.02
			Polytomous	0.94	0.12	1.01	0.15	-0.06	1.01	0.13	1.02	0.15	-0.01
	1.0	One	Dichotomous	0.61	0.10	1.02	0.15	-0.41	0.98	0.15	1.01	0.15	-0.02
			Polytomous	0.44	0.17	1.02	0.15	-0.57	0.66	0.18	1.01	0.15	-0.35
		Composite	Dichotomous	0.56	0.10	1.02	0.15	-0.45	1.19	0.20	1.01	0.15	0.19
			Polytomous	0.49	0.21	1.02	0.15	-0.53	0.69	0.26	1.01	0.15	-0.32
		Separate	Dichotomous	1.11	0.20	1.02	0.15	0.10	1.10	0.19	1.01	0.15	0.10
			Polytomous	1.42	0.17	1.02	0.15	0.41	1.47	0.16	1.01	0.15	0.46
None	0.2	One	Dichotomous	0.68	0.10	1.01	0.15	-0.33	1.16	0.17	1.01	0.15	0.15
			Polytomous	0.56	0.18	1.01	0.15	-0.45	1.02	0.18	1.01	0.15	0.01
		Composite	Dichotomous	0.65	0.11	1.01	0.15	-0.36	1.29	0.22	1.01	0.15	0.28
			Polytomous	0.60	0.21	1.01	0.15	-0.41	1.11	0.26	1.01	0.15	0.10
		Separate	Dichotomous	1.17	0.21	1.01	0.15	0.16	1.19	0.21	1.01	0.15	0.18
			Polytomous	1.81	0.19	1.01	0.15	0.81	1.87	0.19	1.01	0.15	0.86
	0.5	One	Dichotomous	0.82	0.11	1.00	0.10	-0.18	0.87	0.12	1.00	0.10	-0.13
			Polytomous	0.66	0.17	1.00	0.10	-0.34	0.70	0.18	1.00	0.10	-0.30
		Composite	Dichotomous	0.84	0.12	1.00	0.10	-0.17	0.89	0.14	1.00	0.10	-0.11
			Polytomous	0.73	0.18	1.00	0.10	-0.28	0.82	0.19	1.00	0.10	-0.18
		Separate	Dichotomous	1.17	0.41	1.00	0.10	0.16	1.03	0.22	1.00	0.10	0.03
			Polytomous	0.95	0.12	1.00	0.10	-0.06	1.00	0.12	1.00	0.10	0.00
	0.7	One	Dichotomous	1.01	0.13	1.00	0.10	0.01	1.03	0.14	1.01	0.10	0.03
			Polytomous	1.16	0.18	1.00	0.10	0.16	1.21	0.17	1.01	0.10	0.20
		Composite	Dichotomous	1.02	0.14	1.00	0.10	0.02	1.05	0.15	1.01	0.10	0.04
			Polytomous	1.23	0.18	1.00	0.10	0.23	1.32	0.17	1.01	0.10	0.32
		Separate	Dichotomous	1.10	0.18	1.00	0.10	0.10	1.11	0.17	1.01	0.10	0.10

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	1.39	0.15	1.00	0.10	0.39	1.45	0.15	1.01	0.10	0.44
	1.0	One	Dichotomous	1.14	0.14	1.00	0.09	0.14	1.14	0.14	1.01	0.10	0.14
			Polytomous	1.67	0.17	1.00	0.09	0.67	1.68	0.18	1.01	0.10	0.68
		Composite	Dichotomous	1.15	0.15	1.00	0.09	0.15	1.16	0.16	1.01	0.10	0.15
			Polytomous	1.74	0.17	1.00	0.09	0.74	1.78	0.16	1.01	0.10	0.78
		Separate	Dichotomous	1.17	0.17	1.00	0.09	0.17	1.18	0.18	1.01	0.10	0.17
			Polytomous	1.85	0.15	1.00	0.09	0.85	1.87	0.15	1.01	0.10	0.87

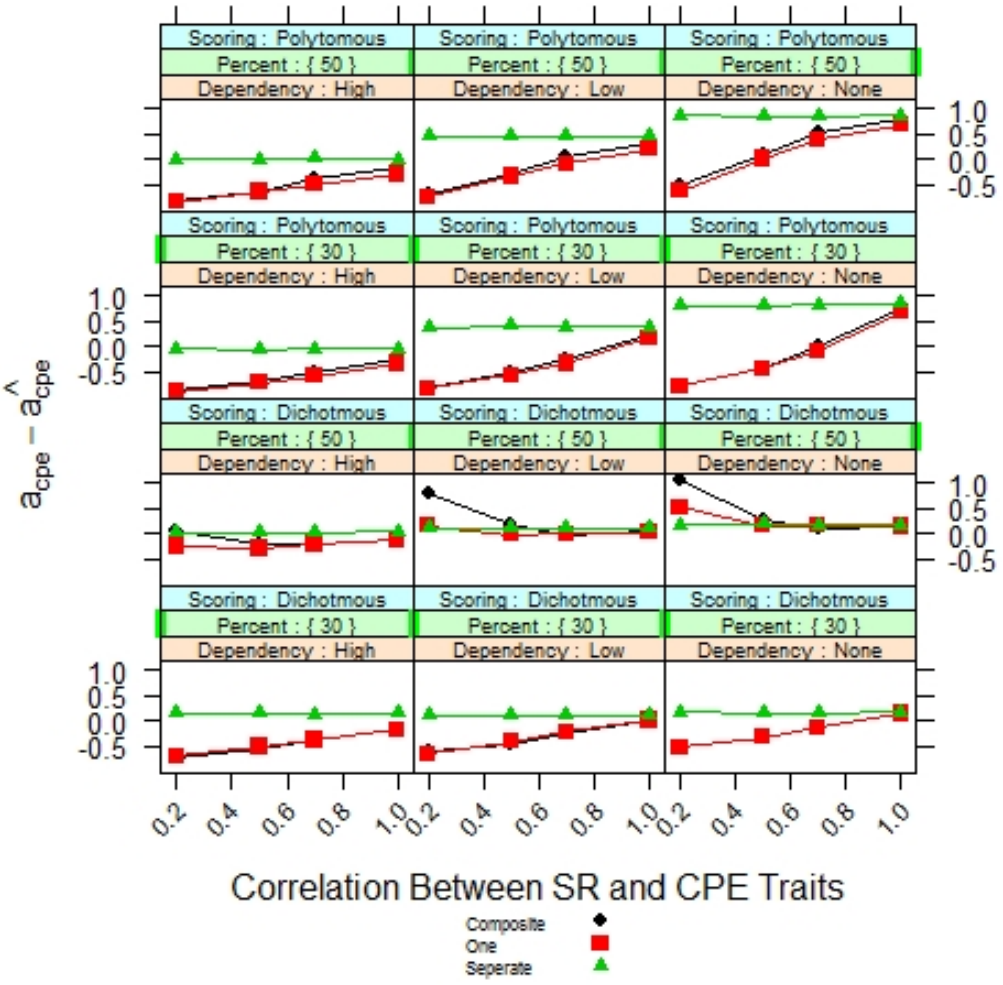

Figure 14: Difference Between Estimated and True CPE "a" Parameters for Sample Size of 1000 and 60 Items

Table 15: Average Estimated and True CPE *a* Parameter for 1000 Test- Takers and 120 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.62	0.10	1.01	0.15	-0.39	0.78	0.12	1.01	0.15	-0.24
			Polytomous	0.43	0.17	1.01	0.15	-0.58	0.51	0.17	1.01	0.15	-0.50
		Composite	Dichotomous	0.60	0.11	1.01	0.15	-0.41	0.74	0.13	1.01	0.15	-0.28
			Polytomous	0.48	0.19	1.01	0.15	-0.53	0.63	0.21	1.01	0.15	-0.38
		Separate	Dichotomous	1.01	0.19	1.01	0.15	-0.01	0.97	0.17	1.01	0.15	-0.04
			Polytomous	1.03	0.14	1.01	0.15	0.02	1.06	0.15	1.01	0.15	0.05
	0.5	One	Dichotomous	0.77	0.11	1.02	0.15	-0.25	1.01	0.15	1.01	0.15	0.00
			Polytomous	0.69	0.17	1.02	0.15	-0.33	0.94	0.17	1.01	0.15	-0.07
		Composite	Dichotomous	0.75	0.12	1.02	0.15	-0.27	0.94	0.15	1.01	0.15	-0.07
			Polytomous	0.77	0.18	1.02	0.15	-0.25	1.10	0.18	1.01	0.15	0.09
		Separate	Dichotomous	1.11	0.19	1.02	0.15	0.09	1.09	0.19	1.01	0.15	0.08
			Polytomous	1.47	0.18	1.02	0.15	0.45	1.52	0.18	1.01	0.15	0.50
	0.7	One	Dichotomous	0.87	0.12	1.01	0.15	-0.14	1.16	0.18	1.01	0.15	0.14
			Polytomous	0.94	0.17	1.01	0.15	-0.07	1.41	0.17	1.01	0.15	0.39
		Composite	Dichotomous	0.85	0.13	1.01	0.15	-0.16	1.05	0.17	1.01	0.15	0.04
			Polytomous	1.03	0.18	1.01	0.15	0.02	1.56	0.17	1.01	0.15	0.55
		Separate	Dichotomous	1.19	0.22	1.01	0.15	0.18	1.18	0.21	1.01	0.15	0.17
			Polytomous	1.91	0.19	1.01	0.15	0.90	1.92	0.20	1.01	0.15	0.91
	1.0	One	Dichotomous	0.33	0.21	1.01	0.15	-0.67	0.67	0.20	1.01	0.15	-0.34
			Polytomous	0.13	0.17	1.01	0.15	-0.88	0.17	0.18	1.01	0.15	-0.84
		Composite	Dichotomous	0.35	0.17	1.01	0.15	-0.66	0.60	0.11	1.01	0.15	-0.41
			Polytomous	0.15	0.21	1.01	0.15	-0.85	0.22	0.26	1.01	0.15	-0.79
		Separate	Dichotomous	1.00	0.19	1.01	0.15	-0.01	0.97	0.18	1.01	0.15	-0.04
			Polytomous	1.03	0.15	1.01	0.15	0.03	1.07	0.15	1.01	0.15	0.06
Low	0.2	One	Dichotomous	0.46	0.23	1.01	0.15	-0.55	1.10	0.19	1.01	0.16	0.09
			Polytomous	0.18	0.17	1.01	0.15	-0.83	0.27	0.17	1.01	0.16	-0.74
		Composite	Dichotomous	0.41	0.14	1.01	0.15	-0.60	2.08	0.31	1.01	0.16	1.07
			Polytomous	0.22	0.21	1.01	0.15	-0.79	0.33	0.28	1.01	0.16	-0.69
		Separate	Dichotomous	1.10	0.20	1.01	0.15	0.09	1.09	0.19	1.01	0.16	0.08
			Polytomous	1.47	0.17	1.01	0.15	0.46	1.52	0.18	1.01	0.16	0.50
	0.5	One	Dichotomous	0.54	0.29	1.01	0.16	-0.48	1.83	0.30	1.01	0.15	0.83
			Polytomous	0.21	0.17	1.01	0.16	-0.80	0.41	0.17	1.01	0.15	-0.60
		Composite	Dichotomous	0.47	0.16	1.01	0.16	-0.54	2.01	0.30	1.01	0.15	1.01
			Polytomous	0.26	0.22	1.01	0.16	-0.75	0.50	0.30	1.01	0.15	-0.51
		Separate	Dichotomous	1.19	0.21	1.01	0.16	0.18	1.17	0.21	1.01	0.15	0.17
			Polytomous	1.88	0.19	1.01	0.16	0.87	1.92	0.19	1.01	0.15	0.91
	0.7	One	Dichotomous	0.47	0.09	1.01	0.15	-0.54	0.72	0.12	1.01	0.15	-0.29
			Polytomous	0.29	0.17	1.01	0.15	-0.72	0.37	0.17	1.01	0.15	-0.64
		Composite	Dichotomous	0.42	0.08	1.01	0.15	-0.58	0.64	0.12	1.01	0.15	-0.37
			Polytomous	0.32	0.20	1.01	0.15	-0.69	0.41	0.24	1.01	0.15	-0.60
		Separate	Dichotomous	1.01	0.20	1.01	0.15	0.00	0.97	0.17	1.01	0.15	-0.04
			Polytomous	1.03	0.13	1.01	0.15	0.02	1.06	0.13	1.01	0.15	0.05
	1.0	One	Dichotomous	0.59	0.09	1.01	0.15	-0.42	0.98	0.14	1.01	0.15	-0.03
			Polytomous	0.45	0.17	1.01	0.15	-0.56	0.67	0.17	1.01	0.15	-0.34
		Composite	Dichotomous	0.52	0.09	1.01	0.15	-0.49	0.88	0.14	1.01	0.15	-0.12
			Polytomous	0.48	0.20	1.01	0.15	-0.53	0.66	0.27	1.01	0.15	-0.35
		Separate	Dichotomous	1.10	0.18	1.01	0.15	0.08	1.09	0.19	1.01	0.15	0.08
			Polytomous	1.46	0.18	1.01	0.15	0.45	1.51	0.18	1.01	0.15	0.50
None	0.2	One	Dichotomous	0.64	0.09	1.01	0.15	-0.37	1.24	0.19	1.01	0.15	0.23
			Polytomous	0.57	0.17	1.01	0.15	-0.44	1.11	0.17	1.01	0.15	0.10
		Composite	Dichotomous	0.56	0.10	1.01	0.15	-0.45	1.22	0.21	1.01	0.15	0.21
			Polytomous	0.59	0.20	1.01	0.15	-0.42	1.25	0.21	1.01	0.15	0.24
		Separate	Dichotomous	1.19	0.21	1.01	0.15	0.18	1.18	0.21	1.01	0.15	0.17
			Polytomous	1.90	0.19	1.01	0.15	0.89	1.93	0.19	1.01	0.15	0.92
	0.5	One	Dichotomous	0.81	0.11	1.00	0.10	-0.20	0.85	0.12	1.01	0.10	-0.15
			Polytomous	0.68	0.16	1.00	0.10	-0.32	0.71	0.18	1.01	0.10	-0.29
		Composite	Dichotomous	0.81	0.12	1.00	0.10	-0.19	0.85	0.13	1.01	0.10	-0.15
			Polytomous	0.75	0.17	1.00	0.10	-0.25	0.85	0.19	1.01	0.10	-0.16
		Separate	Dichotomous	1.01	0.19	1.00	0.10	0.01	0.97	0.15	1.01	0.10	-0.03
			Polytomous	1.03	0.13	1.00	0.10	0.03	1.07	0.13	1.01	0.10	0.06
	0.7	One	Dichotomous	0.99	0.13	1.00	0.10	-0.02	1.02	0.13	1.01	0.10	0.01
			Polytomous	1.15	0.16	1.00	0.10	0.15	1.20	0.17	1.01	0.10	0.20
		Composite	Dichotomous	0.99	0.13	1.00	0.10	-0.01	1.02	0.14	1.01	0.10	0.02
			Polytomous	1.23	0.16	1.00	0.10	0.23	1.34	0.17	1.01	0.10	0.33
		Separate	Dichotomous	1.09	0.15	1.00	0.10	0.09	1.09	0.16	1.01	0.10	0.08

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	1.46	0.15	1.00	0.10	0.46	1.51	0.15	1.01	0.10	0.50
	1.0	One	Dichotomous	1.11	0.14	1.01	0.10	0.11	1.12	0.15	1.01	0.10	0.11
			Polytomous	1.64	0.17	1.01	0.10	0.63	1.65	0.16	1.01	0.10	0.64
		Composite	Dichotomous	1.11	0.15	1.01	0.10	0.11	1.12	0.16	1.01	0.10	0.12
			Polytomous	1.72	0.16	1.01	0.10	0.71	1.79	0.15	1.01	0.10	0.79
		Separate	Dichotomous	1.17	0.17	1.01	0.10	0.16	1.17	0.17	1.01	0.10	0.16
			Polytomous	1.88	0.16	1.01	0.10	0.88	1.94	0.15	1.01	0.10	0.93

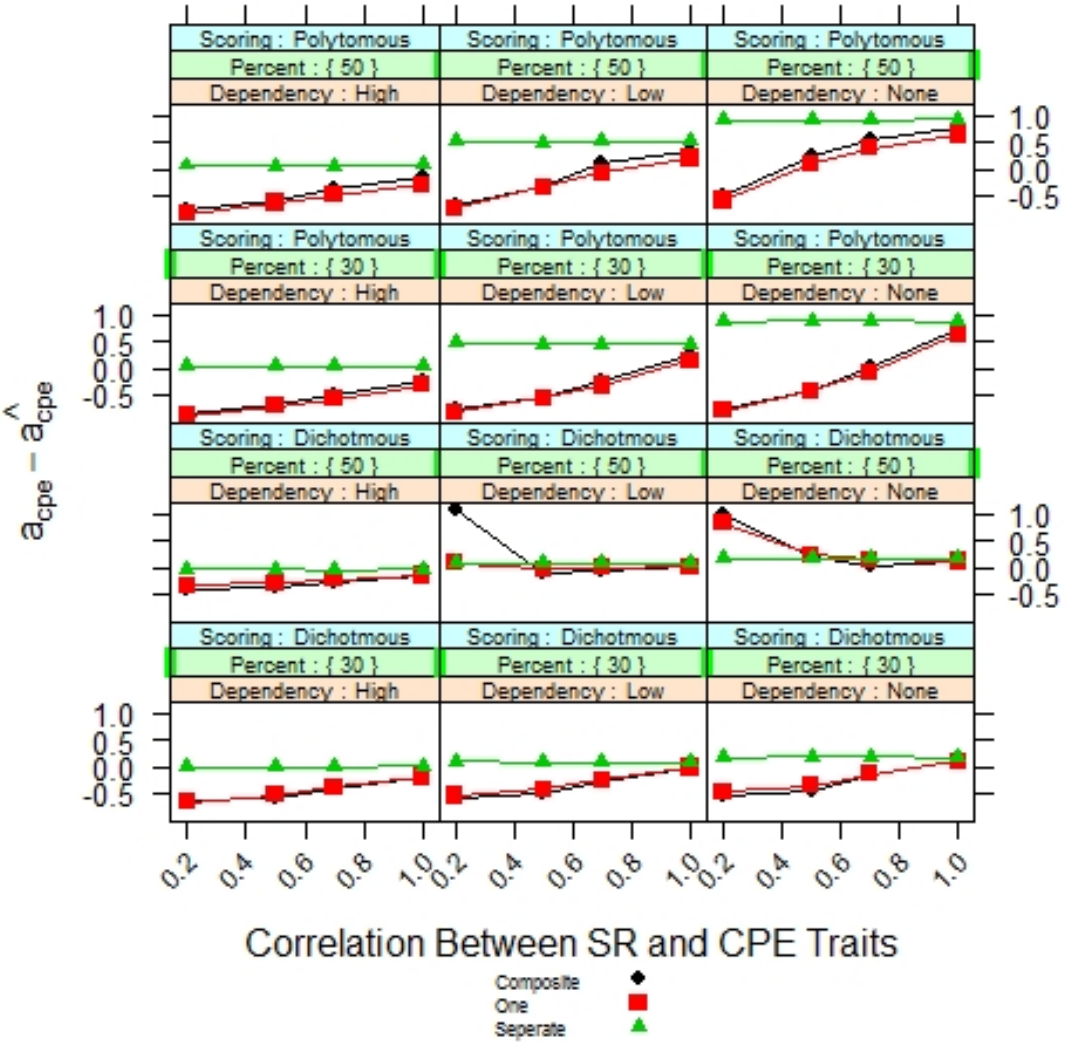

Figure 15: Difference Between Estimated and True CPE "a" Parameters for Sample Size of 1000 and 120 Items

Table 16: Average Estimated and True CPE *a* Parameter for 3000 Test- Takers and 60 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.57	0.07	1.01	0.15	-0.43	0.74	0.10	1.01	0.15	-0.27
			Polytomous	0.42	0.13	1.01	0.15	-0.58	0.50	0.13	1.01	0.15	-0.51
		Composite	Dichotomous	0.56	0.07	1.01	0.15	-0.45	0.73	0.10	1.01	0.15	-0.28
			Polytomous	0.48	0.15	1.01	0.15	-0.52	0.61	0.17	1.01	0.15	-0.40
		Separate	Dichotomous	1.15	0.39	1.01	0.15	0.14	0.97	0.18	1.01	0.15	-0.04
			Polytomous	0.95	0.12	1.01	0.15	-0.06	1.00	0.12	1.01	0.15	-0.01
	0.5	One	Dichotomous	0.73	0.09	1.02	0.15	-0.28	0.95	0.13	1.01	0.15	-0.06
			Polytomous	0.69	0.13	1.02	0.15	-0.32	0.91	0.13	1.01	0.15	-0.10
		Composite	Dichotomous	0.72	0.09	1.02	0.15	-0.29	0.92	0.13	1.01	0.15	-0.09
			Polytomous	0.77	0.15	1.02	0.15	-0.24	1.06	0.15	1.01	0.15	0.05
		Separate	Dichotomous	1.05	0.16	1.02	0.15	0.03	1.05	0.16	1.01	0.15	0.04
			Polytomous	1.41	0.15	1.02	0.15	0.39	1.44	0.16	1.01	0.15	0.43
	0.7	One	Dichotomous	0.83	0.09	1.02	0.15	-0.19	1.11	0.14	1.01	0.15	0.09
			Polytomous	0.93	0.13	1.02	0.15	-0.08	1.43	0.13	1.01	0.15	0.41
		Composite	Dichotomous	0.81	0.10	1.02	0.15	-0.20	1.05	0.14	1.01	0.15	0.03
			Polytomous	1.02	0.15	1.02	0.15	0.01	1.58	0.12	1.01	0.15	0.56
		Separate	Dichotomous	1.11	0.17	1.02	0.15	0.09	1.13	0.17	1.01	0.15	0.12
			Polytomous	1.85	0.14	1.02	0.15	0.84	1.90	0.16	1.01	0.15	0.88
	1.0	One	Dichotomous	0.41	0.18	1.01	0.15	-0.60	0.82	0.20	1.01	0.15	-0.19
			Polytomous	0.12	0.13	1.01	0.15	-0.90	0.15	0.14	1.01	0.15	-0.86
		Composite	Dichotomous	0.44	0.12	1.01	0.15	-0.57	1.02	0.16	1.01	0.15	0.01
			Polytomous	0.14	0.17	1.01	0.15	-0.87	0.22	0.23	1.01	0.15	-0.80
		Separate	Dichotomous	1.15	0.40	1.01	0.15	0.13	0.97	0.16	1.01	0.15	-0.05
			Polytomous	0.95	0.11	1.01	0.15	-0.07	1.00	0.12	1.01	0.15	-0.01
Low	0.2	One	Dichotomous	0.58	0.23	1.00	0.15	-0.42	0.95	0.14	1.01	0.15	-0.06
			Polytomous	0.17	0.12	1.00	0.15	-0.83	0.26	0.13	1.01	0.15	-0.76
		Composite	Dichotomous	0.52	0.15	1.00	0.15	-0.48	1.91	0.21	1.01	0.15	0.90
			Polytomous	0.21	0.16	1.00	0.15	-0.79	0.33	0.25	1.01	0.15	-0.68
		Separate	Dichotomous	1.04	0.16	1.00	0.15	0.04	1.05	0.16	1.01	0.15	0.04
			Polytomous	1.41	0.14	1.00	0.15	0.41	1.46	0.16	1.01	0.15	0.45
	0.5	One	Dichotomous	0.73	0.27	1.00	0.15	-0.27	1.99	0.30	1.01	0.15	0.97
			Polytomous	0.20	0.13	1.00	0.15	-0.80	0.36	0.13	1.01	0.15	-0.65
		Composite	Dichotomous	0.71	0.24	1.00	0.15	-0.29	2.33	0.26	1.01	0.15	1.32
			Polytomous	0.26	0.18	1.00	0.15	-0.74	0.49	0.25	1.01	0.15	-0.53
		Separate	Dichotomous	1.10	0.17	1.00	0.15	0.10	1.13	0.18	1.01	0.15	0.12
			Polytomous	1.84	0.16	1.00	0.15	0.84	1.89	0.17	1.01	0.15	0.87
	0.7	One	Dichotomous	0.44	0.06	1.01	0.15	-0.56	0.64	0.08	1.01	0.15	-0.37
			Polytomous	0.29	0.13	1.01	0.15	-0.72	0.36	0.13	1.01	0.15	-0.65
		Composite	Dichotomous	0.39	0.05	1.01	0.15	-0.62	0.55	0.08	1.01	0.15	-0.46
			Polytomous	0.34	0.16	1.01	0.15	-0.67	0.37	0.22	1.01	0.15	-0.64
		Separate	Dichotomous	1.18	0.46	1.01	0.15	0.17	0.97	0.18	1.01	0.15	-0.04
			Polytomous	0.96	0.11	1.01	0.15	-0.05	1.00	0.12	1.01	0.15	0.00
	1.0	One	Dichotomous	0.55	0.06	1.01	0.14	-0.46	0.91	0.12	1.01	0.15	-0.10
			Polytomous	0.44	0.13	1.01	0.14	-0.57	0.66	0.13	1.01	0.15	-0.35
		Composite	Dichotomous	0.49	0.07	1.01	0.14	-0.52	0.92	0.13	1.01	0.15	-0.09
			Polytomous	0.48	0.17	1.01	0.14	-0.52	0.62	0.29	1.01	0.15	-0.39
		Separate	Dichotomous	1.05	0.16	1.01	0.14	0.04	1.05	0.16	1.01	0.15	0.04
			Polytomous	1.40	0.15	1.01	0.14	0.39	1.46	0.16	1.01	0.15	0.45
None	0.2	One	Dichotomous	0.63	0.06	1.01	0.16	-0.38	1.15	0.14	1.01	0.14	0.14
			Polytomous	0.57	0.13	1.01	0.16	-0.44	1.09	0.13	1.01	0.14	0.08
		Composite	Dichotomous	0.56	0.07	1.01	0.16	-0.45	1.32	0.19	1.01	0.14	0.32
			Polytomous	0.61	0.16	1.01	0.16	-0.40	1.20	0.22	1.01	0.14	0.19
		Separate	Dichotomous	1.10	0.18	1.01	0.16	0.09	1.13	0.17	1.01	0.14	0.12
			Polytomous	1.82	0.16	1.01	0.16	0.81	1.87	0.17	1.01	0.14	0.87
	0.5	One	Dichotomous	0.79	0.09	1.01	0.10	-0.22	0.82	0.09	1.00	0.10	-0.18
			Polytomous	0.69	0.13	1.01	0.10	-0.33	0.71	0.13	1.00	0.10	-0.29
		Composite	Dichotomous	0.80	0.09	1.01	0.10	-0.21	0.84	0.10	1.00	0.10	-0.16
			Polytomous	0.75	0.14	1.01	0.10	-0.26	0.83	0.15	1.00	0.10	-0.18
		Separate	Dichotomous	1.12	0.33	1.01	0.10	0.11	0.96	0.14	1.00	0.10	-0.05
			Polytomous	0.95	0.08	1.01	0.10	-0.06	1.00	0.10	1.00	0.10	-0.01
	0.7	One	Dichotomous	0.97	0.11	1.00	0.10	-0.03	0.99	0.11	1.01	0.10	-0.02
			Polytomous	1.20	0.13	1.00	0.10	0.19	1.23	0.13	1.01	0.10	0.22
		Composite	Dichotomous	0.99	0.12	1.00	0.10	-0.02	1.01	0.12	1.01	0.10	0.00
			Polytomous	1.26	0.13	1.00	0.10	0.26	1.35	0.13	1.01	0.10	0.34
		Separate	Dichotomous	1.05	0.13	1.00	0.10	0.04	1.05	0.13	1.01	0.10	0.05

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	1.42	0.13	1.00	0.10	0.42	1.47	0.13	1.01	0.10	0.47
	1.0	One	Dichotomous	1.10	0.11	1.01	0.10	0.09	1.10	0.12	1.01	0.10	0.09
			Polytomous	1.72	0.13	1.01	0.10	0.72	1.71	0.13	1.01	0.10	0.71
		Composite	Dichotomous	1.11	0.12	1.01	0.10	0.11	1.12	0.13	1.01	0.10	0.11
			Polytomous	1.77	0.13	1.01	0.10	0.77	1.82	0.13	1.01	0.10	0.81
		Separate	Dichotomous	1.10	0.12	1.01	0.10	0.10	1.12	0.13	1.01	0.10	0.12
			Polytomous	1.84	0.12	1.01	0.10	0.83	1.89	0.13	1.01	0.10	0.89

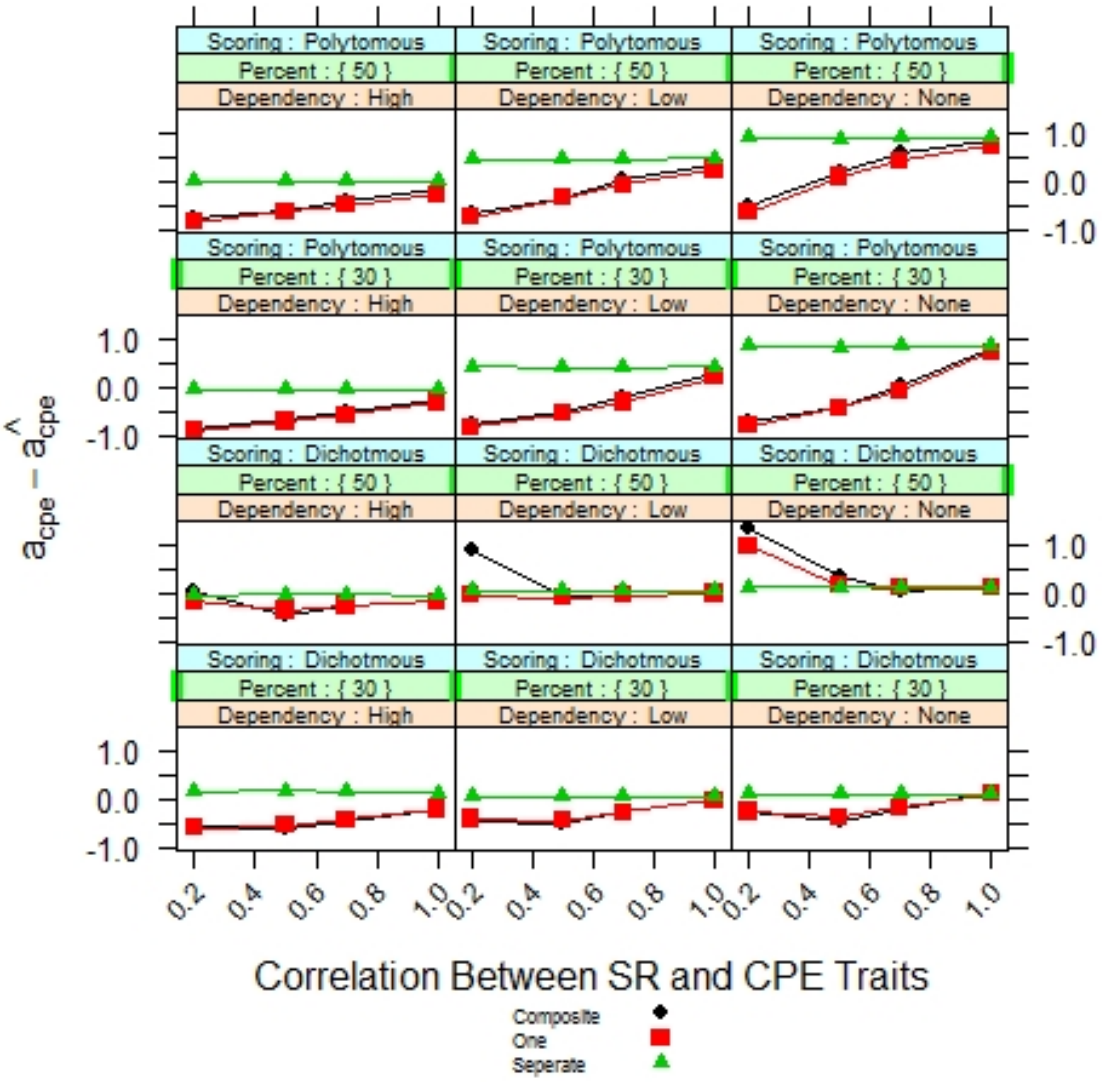

Figure 16: Difference Between Estimated and True CPE "a" Parameters for Sample Size of 3000 and 60 Items

Table 17: Average Estimated and True CPE *a* Parameter for 3000 Test- Takers and 120 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.57	0.07	1.02	0.16	-0.45	0.73	0.09	1.02	0.15	-0.28
			Polytomous	0.43	0.13	1.02	0.16	-0.59	0.52	0.13	1.02	0.15	-0.50
		Composite	Dichotomous	0.55	0.08	1.02	0.16	-0.47	0.69	0.09	1.02	0.15	-0.33
			Polytomous	0.49	0.15	1.02	0.16	-0.53	0.64	0.17	1.02	0.15	-0.38
		Separate	Dichotomous	0.95	0.15	1.02	0.16	-0.08	0.91	0.13	1.02	0.15	-0.10
			Polytomous	1.02	0.12	1.02	0.16	0.00	1.06	0.12	1.02	0.15	0.05
	0.5	One	Dichotomous	0.72	0.09	1.01	0.15	-0.30	0.97	0.13	1.01	0.15	-0.04
			Polytomous	0.70	0.13	1.01	0.15	-0.32	0.96	0.13	1.01	0.15	-0.06
		Composite	Dichotomous	0.70	0.09	1.01	0.15	-0.32	0.89	0.12	1.01	0.15	-0.12
			Polytomous	0.77	0.15	1.01	0.15	-0.24	1.12	0.15	1.01	0.15	0.10
		Separate	Dichotomous	1.05	0.16	1.01	0.15	0.04	1.05	0.15	1.01	0.15	0.03
			Polytomous	1.48	0.17	1.01	0.15	0.46	1.52	0.17	1.01	0.15	0.50
	0.7	One	Dichotomous	0.83	0.09	1.01	0.15	-0.19	1.12	0.15	1.01	0.15	0.11
			Polytomous	0.96	0.13	1.01	0.15	-0.05	1.46	0.13	1.01	0.15	0.44
		Composite	Dichotomous	0.81	0.09	1.01	0.15	-0.21	1.01	0.14	1.01	0.15	0.00
			Polytomous	1.05	0.14	1.01	0.15	0.04	1.59	0.12	1.01	0.15	0.58
		Separate	Dichotomous	1.14	0.17	1.01	0.15	0.13	1.15	0.18	1.01	0.15	0.13
			Polytomous	1.92	0.18	1.01	0.15	0.90	1.94	0.17	1.01	0.15	0.93
	1.0	One	Dichotomous	0.28	0.12	1.01	0.15	-0.73	0.56	0.09	1.01	0.15	-0.45
			Polytomous	0.12	0.13	1.01	0.15	-0.89	0.16	0.13	1.01	0.15	-0.85
		Composite	Dichotomous	0.29	0.07	1.01	0.15	-0.72	0.41	0.05	1.01	0.15	-0.60
			Polytomous	0.15	0.17	1.01	0.15	-0.86	0.22	0.23	1.01	0.15	-0.79
		Separate	Dichotomous	0.94	0.15	1.01	0.15	-0.06	0.91	0.13	1.01	0.15	-0.10
			Polytomous	1.03	0.11	1.01	0.15	0.02	1.07	0.13	1.01	0.15	0.06
Low	0.2	One	Dichotomous	0.43	0.16	1.01	0.15	-0.58	1.29	0.20	1.02	0.15	0.28
			Polytomous	0.17	0.12	1.01	0.15	-0.84	0.27	0.13	1.02	0.15	-0.75
		Composite	Dichotomous	0.46	0.13	1.01	0.15	-0.55	1.76	0.19	1.02	0.15	0.74
			Polytomous	0.22	0.17	1.01	0.15	-0.79	0.34	0.24	1.02	0.15	-0.68
		Separate	Dichotomous	1.04	0.16	1.01	0.15	0.03	1.05	0.16	1.02	0.15	0.03
			Polytomous	1.46	0.17	1.01	0.15	0.45	1.51	0.17	1.02	0.15	0.49
	0.5	One	Dichotomous	0.53	0.19	1.00	0.15	-0.48	2.17	0.29	1.01	0.15	1.16
			Polytomous	0.21	0.13	1.00	0.15	-0.79	0.40	0.13	1.01	0.15	-0.61
		Composite	Dichotomous	0.36	0.06	1.00	0.15	-0.65	2.25	0.25	1.01	0.15	1.24
			Polytomous	0.28	0.19	1.00	0.15	-0.72	0.48	0.31	1.01	0.15	-0.53
		Separate	Dichotomous	1.13	0.17	1.00	0.15	0.12	1.14	0.18	1.01	0.15	0.13
			Polytomous	1.91	0.17	1.00	0.15	0.90	1.96	0.18	1.01	0.15	0.94
	0.7	One	Dichotomous	0.43	0.06	1.01	0.15	-0.58	0.64	0.08	1.01	0.15	-0.38
			Polytomous	0.29	0.13	1.01	0.15	-0.72	0.37	0.13	1.01	0.15	-0.64
		Composite	Dichotomous	0.38	0.05	1.01	0.15	-0.63	0.47	0.07	1.01	0.15	-0.54
			Polytomous	0.33	0.16	1.01	0.15	-0.68	0.39	0.22	1.01	0.15	-0.62
		Separate	Dichotomous	0.95	0.15	1.01	0.15	-0.05	0.91	0.13	1.01	0.15	-0.10
			Polytomous	1.02	0.12	1.01	0.15	0.01	1.07	0.13	1.01	0.15	0.05
	1.0	One	Dichotomous	0.54	0.06	1.01	0.15	-0.48	0.97	0.13	1.01	0.15	-0.04
			Polytomous	0.45	0.13	1.01	0.15	-0.56	0.68	0.13	1.01	0.15	-0.33
		Composite	Dichotomous	0.46	0.06	1.01	0.15	-0.55	0.94	0.14	1.01	0.15	-0.07
			Polytomous	0.49	0.16	1.01	0.15	-0.53	0.65	0.28	1.01	0.15	-0.36
		Separate	Dichotomous	1.05	0.16	1.01	0.15	0.03	1.04	0.16	1.01	0.15	0.03
			Polytomous	1.47	0.16	1.01	0.15	0.46	1.51	0.17	1.01	0.15	0.50
None	0.2	One	Dichotomous	0.62	0.06	1.01	0.15	-0.39	1.23	0.16	1.01	0.15	0.21
			Polytomous	0.59	0.13	1.01	0.15	-0.42	1.20	0.13	1.01	0.15	0.19
		Composite	Dichotomous	0.54	0.08	1.01	0.15	-0.47	1.19	0.17	1.01	0.15	0.17
			Polytomous	0.62	0.16	1.01	0.15	-0.39	1.26	0.23	1.01	0.15	0.24
		Separate	Dichotomous	1.14	0.18	1.01	0.15	0.13	1.14	0.17	1.01	0.15	0.12
			Polytomous	1.93	0.19	1.01	0.15	0.92	1.92	0.17	1.01	0.15	0.91
	0.5	One	Dichotomous	0.76	0.09	1.01	0.10	-0.25	0.80	0.09	1.00	0.10	-0.20
			Polytomous	0.68	0.13	1.01	0.10	-0.32	0.71	0.13	1.00	0.10	-0.29
		Composite	Dichotomous	0.77	0.09	1.01	0.10	-0.24	0.81	0.09	1.00	0.10	-0.20
			Polytomous	0.76	0.14	1.01	0.10	-0.25	0.84	0.15	1.00	0.10	-0.16
		Separate	Dichotomous	0.94	0.13	1.01	0.10	-0.06	0.91	0.10	1.00	0.10	-0.10
			Polytomous	1.03	0.11	1.01	0.10	0.02	1.05	0.10	1.00	0.10	0.05
	0.7	One	Dichotomous	0.96	0.11	1.01	0.10	-0.05	0.98	0.11	1.01	0.10	-0.03
			Polytomous	1.18	0.13	1.01	0.10	0.18	1.22	0.13	1.01	0.10	0.22
		Composite	Dichotomous	0.96	0.11	1.01	0.10	-0.04	0.99	0.11	1.01	0.10	-0.02
			Polytomous	1.26	0.13	1.01	0.10	0.25	1.36	0.14	1.01	0.10	0.35
		Separate	Dichotomous	1.04	0.12	1.01	0.10	0.04	1.04	0.12	1.01	0.10	0.03

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	1.47	0.13	1.01	0.10	0.47	1.53	0.14	1.01	0.10	0.52
	1.0	One	Dichotomous	1.09	0.12	1.01	0.10	0.09	1.09	0.12	1.00	0.10	0.08
			Polytomous	1.70	0.13	1.01	0.10	0.69	1.69	0.13	1.00	0.10	0.69
		Composite	Dichotomous	1.09	0.12	1.01	0.10	0.09	1.10	0.12	1.00	0.10	0.09
			Polytomous	1.76	0.12	1.01	0.10	0.76	1.81	0.13	1.00	0.10	0.81
		Separate	Dichotomous	1.13	0.13	1.01	0.10	0.13	1.13	0.13	1.00	0.10	0.13
			Polytomous	1.91	0.12	1.01	0.10	0.90	1.94	0.13	1.00	0.10	0.93

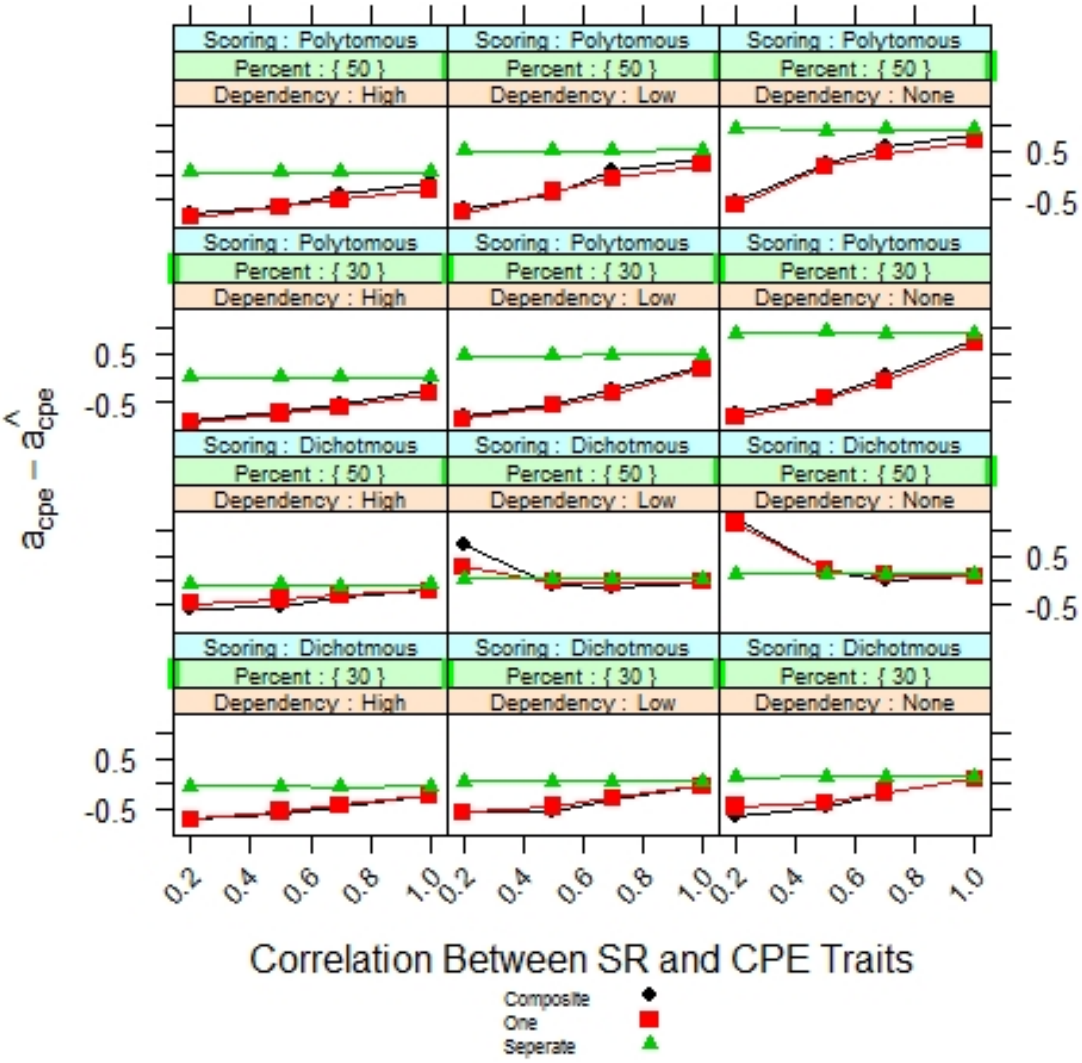

Figure 17: Difference Between Estimated and True CPE "a" Parameters for Sample Size of 3000 and 120 Items

Table 18: Average Estimated and True Selected Response *b* Parameter for 1000 Test-Takers and 60 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.06	0.71	0.00	0.73	0.06	0.05	0.72	0.02	0.75	0.04
			Polytomous	0.06	0.71	0.00	0.73	0.06	0.05	0.72	0.02	0.75	0.04
		Composite	Dichotomous	0.09	0.68	0.00	0.73	0.09	0.13	0.69	0.02	0.75	0.11
			Polytomous	0.06	0.60	0.00	0.73	0.06	0.07	0.54	0.02	0.75	0.05
		Separate	Dichotomous	0.06	0.18	0.00	0.73	0.06	0.05	0.18	0.02	0.75	0.04
			Polytomous	0.06	0.18	0.00	0.73	0.06	0.05	0.18	0.02	0.75	0.04
	0.5	One	Dichotomous	0.05	0.70	0.01	0.73	0.03	0.09	0.70	-0.02	0.73	0.10
			Polytomous	0.05	0.70	0.01	0.73	0.03	0.09	0.70	-0.02	0.73	0.10
		Composite	Dichotomous	0.08	0.67	0.01	0.73	0.06	0.13	0.73	-0.02	0.73	0.14
			Polytomous	0.05	0.60	0.01	0.73	0.04	0.12	0.57	-0.02	0.73	0.14
		Separate	Dichotomous	0.05	0.18	0.01	0.73	0.03	0.09	0.17	-0.02	0.73	0.10
			Polytomous	0.05	0.18	0.01	0.73	0.03	0.09	0.17	-0.02	0.73	0.10
	0.7	One	Dichotomous	0.05	0.71	0.00	0.75	0.04	0.04	0.71	0.03	0.74	0.01
			Polytomous	0.05	0.71	0.00	0.75	0.04	0.04	0.71	0.03	0.74	0.01
		Composite	Dichotomous	0.07	0.70	0.00	0.75	0.07	0.06	0.77	0.03	0.74	0.03
			Polytomous	0.06	0.63	0.00	0.75	0.05	0.08	0.64	0.03	0.74	0.05
		Separate	Dichotomous	0.05	0.18	0.00	0.75	0.04	0.04	0.18	0.03	0.74	0.01
			Polytomous	0.05	0.18	0.00	0.75	0.04	0.04	0.18	0.03	0.74	0.01
	1.0	One	Dichotomous	0.05	0.72	0.00	0.74	0.05	0.05	0.73	0.02	0.75	0.03
			Polytomous	0.05	0.72	0.00	0.74	0.05	0.05	0.73	0.02	0.75	0.03
		Composite	Dichotomous	0.12	0.62	0.00	0.74	0.12	-0.05	1.13	0.02	0.75	-0.07
			Polytomous	0.05	0.57	0.00	0.74	0.05	0.09	0.47	0.02	0.75	0.07
		Separate	Dichotomous	0.05	0.17	0.00	0.74	0.05	0.05	0.17	0.02	0.75	0.03
			Polytomous	0.05	0.17	0.00	0.74	0.05	0.05	0.17	0.02	0.75	0.03
Low	0.2	One	Dichotomous	0.06	0.72	0.00	0.74	0.06	0.07	0.70	0.01	0.73	0.06
			Polytomous	0.06	0.72	0.00	0.74	0.06	0.07	0.70	0.01	0.73	0.06
		Composite	Dichotomous	0.17	0.63	0.00	0.74	0.18	-0.27	1.61	0.01	0.73	-0.27
			Polytomous	0.08	0.57	0.00	0.74	0.08	0.19	0.36	0.01	0.73	0.19
		Separate	Dichotomous	0.06	0.18	0.00	0.74	0.06	0.07	0.18	0.01	0.73	0.06
			Polytomous	0.06	0.18	0.00	0.74	0.06	0.07	0.18	0.01	0.73	0.06
	0.5	One	Dichotomous	0.04	0.71	0.01	0.74	0.03	0.08	0.72	-0.01	0.74	0.09
			Polytomous	0.04	0.71	0.01	0.74	0.03	0.08	0.72	-0.01	0.74	0.09
		Composite	Dichotomous	0.15	0.60	0.01	0.74	0.14	-0.24	1.78	-0.01	0.74	-0.23
			Polytomous	0.07	0.54	0.01	0.74	0.06	0.25	0.50	-0.01	0.74	0.26
		Separate	Dichotomous	0.04	0.17	0.01	0.74	0.03	0.08	0.18	-0.01	0.74	0.09
			Polytomous	0.04	0.17	0.01	0.74	0.03	0.08	0.18	-0.01	0.74	0.09
	0.7	One	Dichotomous	0.05	0.73	0.01	0.75	0.05	0.09	0.70	-0.03	0.74	0.12
			Polytomous	0.05	0.73	0.01	0.75	0.05	0.09	0.70	-0.03	0.74	0.12
		Composite	Dichotomous	0.12	0.67	0.01	0.75	0.12	0.18	0.77	-0.03	0.74	0.21
			Polytomous	0.06	0.60	0.01	0.75	0.06	0.14	0.51	-0.03	0.74	0.17
		Separate	Dichotomous	0.05	0.17	0.01	0.75	0.05	0.09	0.17	-0.03	0.74	0.12
			Polytomous	0.05	0.17	0.01	0.75	0.05	0.09	0.17	-0.03	0.74	0.12
	1.0	One	Dichotomous	0.04	0.74	0.02	0.75	0.01	0.07	0.70	0.00	0.73	0.06
			Polytomous	0.04	0.74	0.02	0.75	0.01	0.07	0.70	0.00	0.73	0.06
		Composite	Dichotomous	0.10	0.70	0.02	0.75	0.08	0.01	0.97	0.00	0.73	0.00
			Polytomous	0.06	0.61	0.02	0.75	0.04	0.16	0.54	0.00	0.73	0.15
		Separate	Dichotomous	0.04	0.17	0.02	0.75	0.01	0.07	0.18	0.00	0.73	0.06
			Polytomous	0.04	0.17	0.02	0.75	0.01	0.07	0.18	0.00	0.73	0.06
None	0.2	One	Dichotomous	0.04	0.72	0.02	0.75	0.02	0.05	0.73	0.02	0.75	0.04
			Polytomous	0.04	0.72	0.02	0.75	0.02	0.05	0.73	0.02	0.75	0.04
		Composite	Dichotomous	0.10	0.68	0.02	0.75	0.08	-0.04	1.06	0.02	0.75	-0.06
			Polytomous	0.07	0.60	0.02	0.75	0.05	0.13	0.66	0.02	0.75	0.12
		Separate	Dichotomous	0.04	0.18	0.02	0.75	0.02	0.05	0.18	0.02	0.75	0.04
			Polytomous	0.04	0.18	0.02	0.75	0.02	0.05	0.18	0.02	0.75	0.04
	0.5	One	Dichotomous	0.05	0.72	0.01	0.74	0.04	0.04	0.73	0.03	0.75	0.01
			Polytomous	0.05	0.72	0.01	0.74	0.04	0.04	0.73	0.03	0.75	0.01
		Composite	Dichotomous	0.06	0.72	0.01	0.74	0.05	0.06	0.72	0.03	0.75	0.03
			Polytomous	0.05	0.65	0.01	0.74	0.04	0.03	0.62	0.03	0.75	0.00
		Separate	Dichotomous	0.05	0.17	0.01	0.74	0.04	0.04	0.18	0.03	0.75	0.01
			Polytomous	0.05	0.17	0.01	0.74	0.04	0.04	0.18	0.03	0.75	0.01
	0.7	One	Dichotomous	0.06	0.71	0.00	0.72	0.06	0.05	0.71	0.02	0.74	0.03
			Polytomous	0.06	0.71	0.00	0.72	0.06	0.05	0.71	0.02	0.74	0.03
		Composite	Dichotomous	0.06	0.71	0.00	0.72	0.06	0.05	0.71	0.02	0.74	0.04
			Polytomous	0.06	0.67	0.00	0.72	0.06	0.04	0.65	0.02	0.74	0.02
		Separate	Dichotomous	0.06	0.18	0.00	0.72	0.06	0.05	0.17	0.02	0.74	0.03

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	0.06	0.18	0.00	0.72	0.06	0.05	0.17	0.02	0.74	0.03
	1.0	One	Dichotomous	0.04	0.72	0.03	0.74	0.01	0.04	0.71	0.03	0.75	0.00
			Polytomous	0.04	0.72	0.03	0.74	0.01	0.04	0.71	0.03	0.75	0.00
		Composite	Dichotomous	0.03	0.72	0.03	0.74	0.00	0.03	0.71	0.03	0.75	-0.01
			Polytomous	0.03	0.69	0.03	0.74	0.00	0.03	0.68	0.03	0.75	-0.01
		Separate	Dichotomous	0.04	0.17	0.03	0.74	0.01	0.04	0.18	0.03	0.75	0.00
			Polytomous	0.04	0.17	0.03	0.74	0.01	0.04	0.18	0.03	0.75	0.00

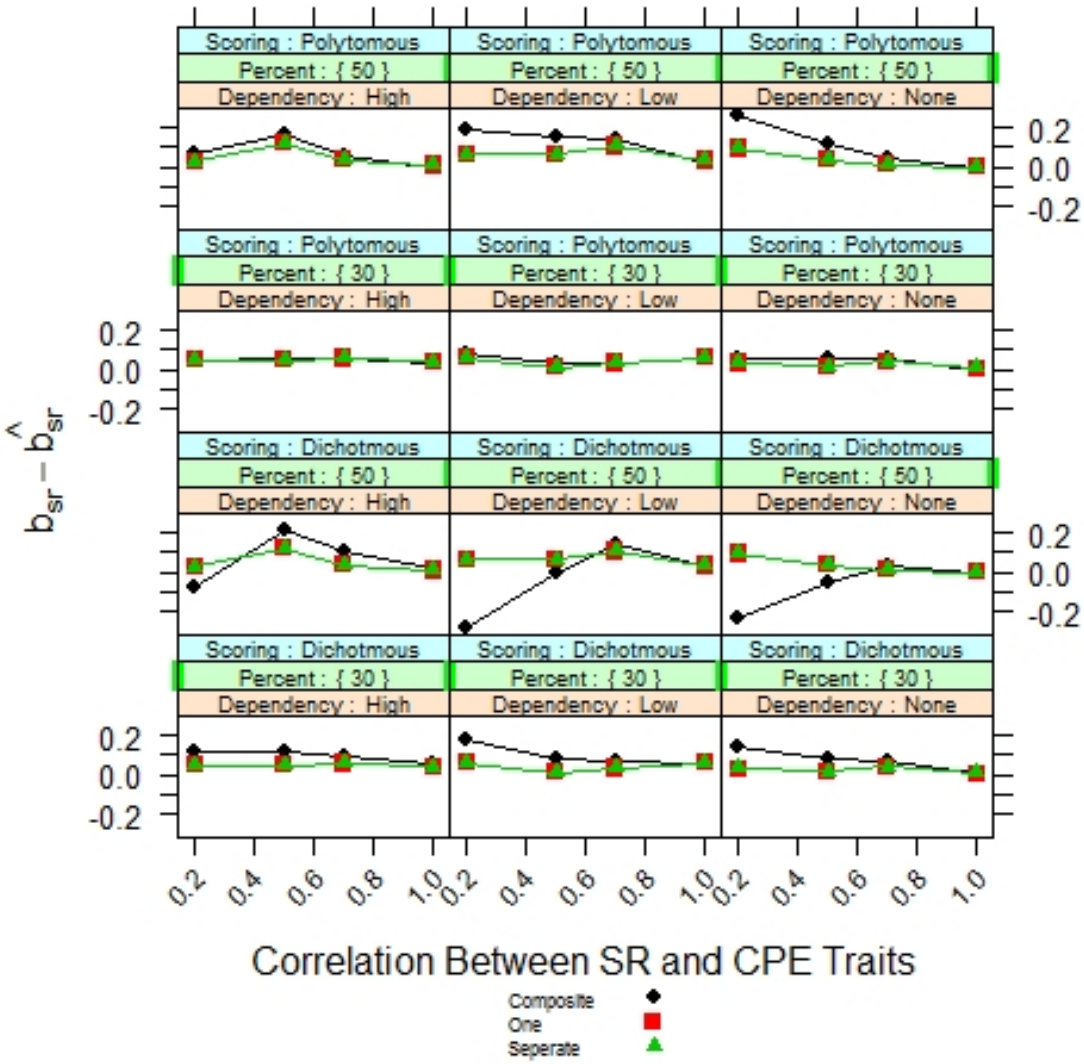

Figure 18: Difference Between Estimated and True SR "b" Parameters for Sample Size of 1000 and 60 Items

Table 19: Average Estimated and True Selected Response *b* Parameter for 1000 Test-Takers and 120 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.03	0.74	0.01	0.75	0.02	0.06	0.71	-0.01	0.73	0.07
			Polytomous	0.03	0.74	0.01	0.75	0.02	0.06	0.71	-0.01	0.73	0.07
		Composite	Dichotomous	0.07	0.71	0.01	0.75	0.07	0.16	0.67	-0.01	0.73	0.17
			Polytomous	0.04	0.62	0.01	0.75	0.03	0.09	0.53	-0.01	0.73	0.09
		Separate	Dichotomous	0.03	0.17	0.01	0.75	0.02	0.06	0.17	-0.01	0.73	0.07
			Polytomous	0.03	0.17	0.01	0.75	0.02	0.06	0.17	-0.01	0.73	0.07
	0.5	One	Dichotomous	0.04	0.73	0.00	0.76	0.04	0.07	0.73	-0.02	0.75	0.09
			Polytomous	0.04	0.73	0.00	0.76	0.04	0.07	0.73	-0.02	0.75	0.09
		Composite	Dichotomous	0.08	0.72	0.00	0.76	0.08	0.14	0.76	-0.02	0.75	0.16
			Polytomous	0.06	0.63	0.00	0.76	0.06	0.11	0.59	-0.02	0.75	0.13
		Separate	Dichotomous	0.04	0.17	0.00	0.76	0.04	0.07	0.17	-0.02	0.75	0.09
			Polytomous	0.04	0.17	0.00	0.76	0.04	0.07	0.17	-0.02	0.75	0.09
	0.7	One	Dichotomous	0.04	0.72	0.00	0.74	0.04	0.05	0.72	0.00	0.74	0.04
			Polytomous	0.04	0.72	0.00	0.74	0.04	0.05	0.72	0.00	0.74	0.04
		Composite	Dichotomous	0.07	0.72	0.00	0.74	0.07	0.10	0.78	0.00	0.74	0.10
			Polytomous	0.05	0.63	0.00	0.74	0.05	0.08	0.64	0.00	0.74	0.08
		Separate	Dichotomous	0.04	0.17	0.00	0.74	0.04	0.05	0.17	0.00	0.74	0.04
			Polytomous	0.04	0.17	0.00	0.74	0.04	0.05	0.17	0.00	0.74	0.04
	1.0	One	Dichotomous	0.05	0.73	-0.01	0.75	0.06	0.06	0.73	-0.01	0.74	0.06
			Polytomous	0.05	0.73	-0.01	0.75	0.06	0.06	0.73	-0.01	0.74	0.06
		Composite	Dichotomous	0.17	0.60	-0.01	0.75	0.18	0.23	0.78	-0.01	0.74	0.24
			Polytomous	0.05	0.56	-0.01	0.75	0.06	0.12	0.41	-0.01	0.74	0.12
		Separate	Dichotomous	0.05	0.17	-0.01	0.75	0.06	0.06	0.18	-0.01	0.74	0.06
			Polytomous	0.05	0.17	-0.01	0.75	0.06	0.06	0.18	-0.01	0.74	0.06
Low	0.2	One	Dichotomous	0.03	0.73	0.01	0.74	0.02	0.04	0.74	0.01	0.76	0.03
			Polytomous	0.03	0.73	0.01	0.74	0.02	0.04	0.74	0.01	0.76	0.03
		Composite	Dichotomous	0.17	0.59	0.01	0.74	0.16	-0.39	1.70	0.01	0.76	-0.40
			Polytomous	0.05	0.54	0.01	0.74	0.05	0.17	0.47	0.01	0.76	0.17
		Separate	Dichotomous	0.03	0.17	0.01	0.74	0.02	0.04	0.17	0.01	0.76	0.03
			Polytomous	0.03	0.17	0.01	0.74	0.02	0.04	0.17	0.01	0.76	0.03
	0.5	One	Dichotomous	0.05	0.75	-0.01	0.76	0.05	0.06	0.71	-0.01	0.74	0.07
			Polytomous	0.05	0.75	-0.01	0.76	0.05	0.06	0.71	-0.01	0.74	0.07
		Composite	Dichotomous	0.18	0.67	-0.01	0.76	0.18	-0.28	1.77	-0.01	0.74	-0.27
			Polytomous	0.09	0.53	-0.01	0.76	0.10	0.23	0.55	-0.01	0.74	0.24
		Separate	Dichotomous	0.05	0.17	-0.01	0.76	0.05	0.06	0.17	-0.01	0.74	0.07
			Polytomous	0.05	0.17	-0.01	0.76	0.05	0.06	0.17	-0.01	0.74	0.07
	0.7	One	Dichotomous	0.04	0.73	-0.01	0.75	0.04	0.03	0.75	0.02	0.76	0.01
			Polytomous	0.04	0.73	-0.01	0.75	0.04	0.03	0.75	0.02	0.76	0.01
		Composite	Dichotomous	0.13	0.66	-0.01	0.75	0.14	0.19	0.75	0.02	0.76	0.17
			Polytomous	0.05	0.58	-0.01	0.75	0.06	0.10	0.53	0.02	0.76	0.08
		Separate	Dichotomous	0.04	0.17	-0.01	0.75	0.04	0.03	0.17	0.02	0.76	0.01
			Polytomous	0.04	0.17	-0.01	0.75	0.04	0.03	0.17	0.02	0.76	0.01
	1.0	One	Dichotomous	0.05	0.74	0.00	0.75	0.05	0.05	0.71	0.00	0.74	0.05
			Polytomous	0.05	0.74	0.00	0.75	0.05	0.05	0.71	0.00	0.74	0.05
		Composite	Dichotomous	0.15	0.69	0.00	0.75	0.15	0.15	0.88	0.00	0.74	0.16
			Polytomous	0.08	0.60	0.00	0.75	0.08	0.16	0.59	0.00	0.74	0.16
		Separate	Dichotomous	0.05	0.17	0.00	0.75	0.05	0.05	0.17	0.00	0.74	0.05
			Polytomous	0.05	0.17	0.00	0.75	0.05	0.05	0.17	0.00	0.74	0.05
None	0.2	One	Dichotomous	0.07	0.73	-0.02	0.75	0.09	0.04	0.71	0.00	0.74	0.04
			Polytomous	0.07	0.73	-0.02	0.75	0.09	0.04	0.71	0.00	0.74	0.04
		Composite	Dichotomous	0.19	0.67	-0.02	0.75	0.21	0.00	1.03	0.00	0.74	0.00
			Polytomous	0.12	0.58	-0.02	0.75	0.15	0.13	0.63	0.00	0.74	0.13
		Separate	Dichotomous	0.07	0.17	-0.02	0.75	0.09	0.04	0.17	0.00	0.74	0.04
			Polytomous	0.07	0.17	-0.02	0.75	0.09	0.04	0.17	0.00	0.74	0.04
	0.5	One	Dichotomous	0.06	0.72	-0.02	0.74	0.08	0.05	0.74	0.00	0.77	0.05
			Polytomous	0.06	0.72	-0.02	0.74	0.08	0.05	0.74	0.00	0.77	0.05
		Composite	Dichotomous	0.08	0.71	-0.02	0.74	0.10	0.08	0.73	0.00	0.77	0.08
			Polytomous	0.06	0.65	-0.02	0.74	0.08	0.05	0.62	0.00	0.77	0.05
		Separate	Dichotomous	0.06	0.16	-0.02	0.74	0.08	0.05	0.18	0.00	0.77	0.05
			Polytomous	0.06	0.16	-0.02	0.74	0.08	0.05	0.18	0.00	0.77	0.05
	0.7	One	Dichotomous	0.06	0.75	-0.02	0.76	0.08	0.05	0.73	-0.01	0.75	0.06
			Polytomous	0.06	0.75	-0.02	0.76	0.08	0.05	0.73	-0.01	0.75	0.06
		Composite	Dichotomous	0.07	0.75	-0.02	0.76	0.09	0.07	0.73	-0.01	0.75	0.07
			Polytomous	0.05	0.70	-0.02	0.76	0.08	0.05	0.66	-0.01	0.75	0.05
		Separate	Dichotomous	0.06	0.16	-0.02	0.76	0.08	0.05	0.17	-0.01	0.75	0.06

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	0.06	0.16	-0.02	0.76	0.08	0.05	0.17	-0.01	0.75	0.06
	1.0	One	Dichotomous	0.03	0.75	0.01	0.75	0.02	0.09	0.72	-0.05	0.74	0.14
			Polytomous	0.03	0.75	0.01	0.75	0.02	0.09	0.72	-0.05	0.74	0.14
		Composite	Dichotomous	0.04	0.76	0.01	0.75	0.02	0.10	0.72	-0.05	0.74	0.15
			Polytomous	0.03	0.72	0.01	0.75	0.02	0.09	0.67	-0.05	0.74	0.14
		Separate	Dichotomous	0.03	0.17	0.01	0.75	0.02	0.09	0.16	-0.05	0.74	0.14
			Polytomous	0.03	0.17	0.01	0.75	0.02	0.09	0.16	-0.05	0.74	0.14

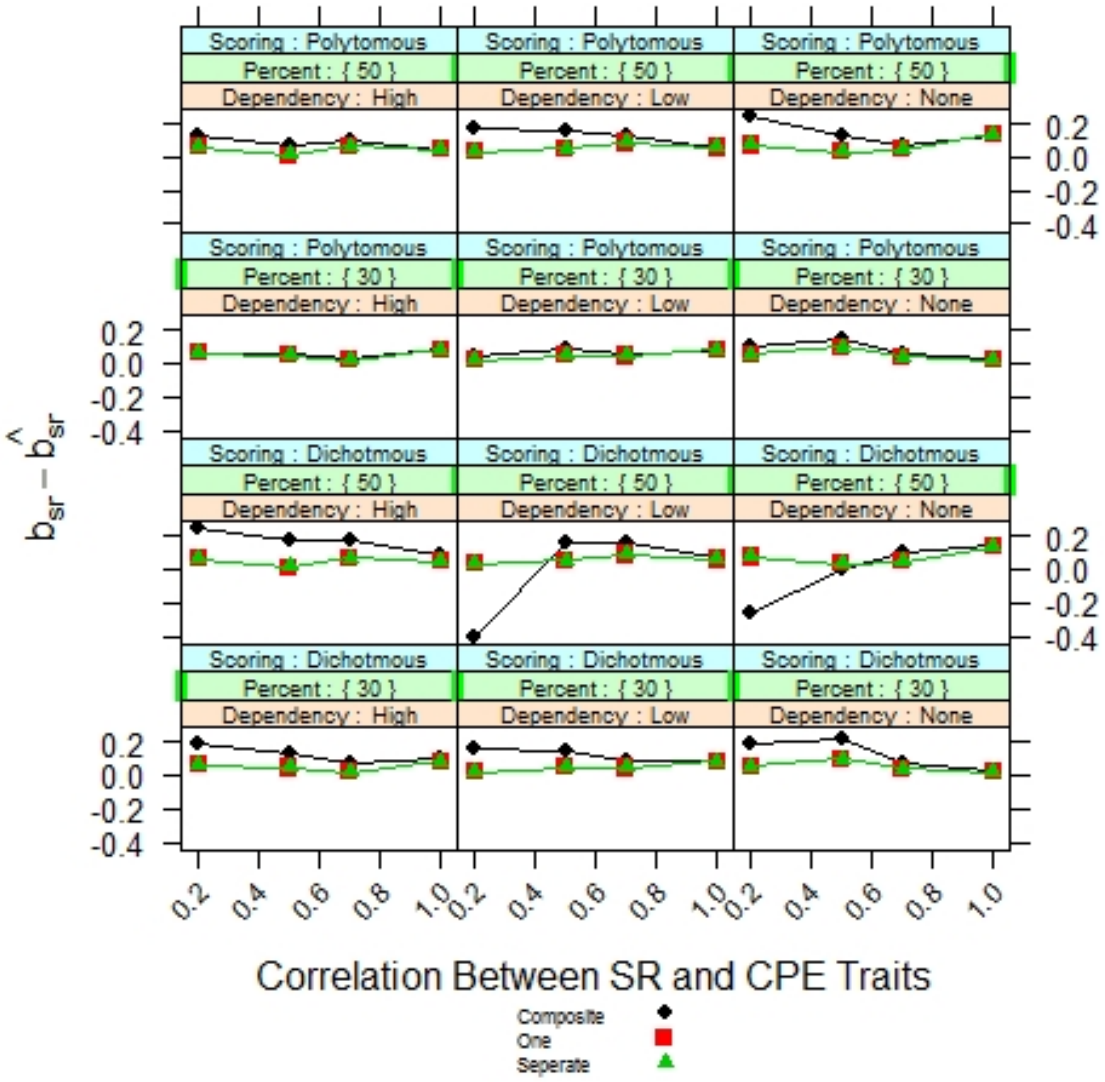

Figure 19: Difference Between Estimated and True SR "b" Parameters for Sample Size of 1000 and 120 Items

Table 20: Average Estimated and True Selected Response *b* Parameter for 3000 Test-Takers and 60 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.08	0.64	0.02	0.75	0.07	0.17	0.61	-0.01	0.73	0.17
			Polytomous	0.03	0.60	0.02	0.75	0.02	0.09	0.50	-0.01	0.73	0.09
		Composite	Dichotomous	0.01	0.71	0.02	0.75	0.00	0.05	0.69	-0.01	0.73	0.06
			Polytomous	0.01	0.71	0.02	0.75	0.00	0.05	0.69	-0.01	0.73	0.06
		Separate	Dichotomous	0.01	0.13	0.02	0.75	0.00	0.05	0.13	-0.01	0.73	0.06
			Polytomous	0.01	0.13	0.02	0.75	0.00	0.05	0.13	-0.01	0.73	0.06
	0.5	One	Dichotomous	0.06	0.69	0.02	0.76	0.04	0.14	0.66	-0.01	0.72	0.15
			Polytomous	0.03	0.63	0.02	0.76	0.01	0.11	0.53	-0.01	0.72	0.13
		Composite	Dichotomous	0.01	0.73	0.02	0.76	-0.02	0.06	0.69	-0.01	0.72	0.07
			Polytomous	0.01	0.73	0.02	0.76	-0.02	0.06	0.69	-0.01	0.72	0.07
		Separate	Dichotomous	0.01	0.13	0.02	0.76	-0.02	0.06	0.13	-0.01	0.72	0.07
			Polytomous	0.01	0.13	0.02	0.76	-0.02	0.06	0.13	-0.01	0.72	0.07
	0.7	One	Dichotomous	0.10	0.67	-0.02	0.76	0.11	0.09	0.72	0.01	0.73	0.08
			Polytomous	0.08	0.62	-0.02	0.76	0.09	0.08	0.61	0.01	0.73	0.07
		Composite	Dichotomous	0.05	0.72	-0.02	0.76	0.06	0.03	0.70	0.01	0.73	0.03
			Polytomous	0.05	0.72	-0.02	0.76	0.06	0.03	0.70	0.01	0.73	0.03
		Separate	Dichotomous	0.05	0.13	-0.02	0.76	0.06	0.03	0.13	0.01	0.73	0.03
			Polytomous	0.05	0.13	-0.02	0.76	0.06	0.03	0.13	0.01	0.73	0.03
	1.0	One	Dichotomous	0.16	0.49	0.01	0.75	0.15	0.04	0.86	0.00	0.74	0.04
			Polytomous	0.01	0.56	0.01	0.75	0.00	0.14	0.33	0.00	0.74	0.14
		Composite	Dichotomous	0.02	0.71	0.01	0.75	0.01	0.04	0.71	0.00	0.74	0.04
			Polytomous	0.02	0.71	0.01	0.75	0.01	0.04	0.71	0.00	0.74	0.04
		Separate	Dichotomous	0.02	0.13	0.01	0.75	0.01	0.04	0.14	0.00	0.74	0.04
			Polytomous	0.02	0.13	0.01	0.75	0.01	0.04	0.14	0.00	0.74	0.04
Low	0.2	One	Dichotomous	0.20	0.49	0.00	0.75	0.20	-0.37	1.71	0.02	0.75	-0.39
			Polytomous	0.05	0.54	0.00	0.75	0.05	0.21	0.36	0.02	0.75	0.19
		Composite	Dichotomous	0.03	0.72	0.00	0.75	0.04	0.03	0.72	0.02	0.75	0.01
			Polytomous	0.03	0.72	0.00	0.75	0.04	0.03	0.72	0.02	0.75	0.01
		Separate	Dichotomous	0.03	0.12	0.00	0.75	0.04	0.03	0.13	0.02	0.75	0.01
			Polytomous	0.03	0.12	0.00	0.75	0.04	0.03	0.13	0.02	0.75	0.01
	0.5	One	Dichotomous	0.12	0.58	0.02	0.74	0.11	-0.46	1.84	0.01	0.74	-0.47
			Polytomous	0.07	0.49	0.02	0.74	0.06	0.22	0.48	0.01	0.74	0.21
		Composite	Dichotomous	0.01	0.72	0.02	0.74	0.00	0.04	0.71	0.01	0.74	0.03
			Polytomous	0.01	0.72	0.02	0.74	0.00	0.04	0.71	0.01	0.74	0.03
		Separate	Dichotomous	0.01	0.13	0.02	0.74	0.00	0.04	0.13	0.01	0.74	0.03
			Polytomous	0.01	0.13	0.02	0.74	0.00	0.04	0.13	0.01	0.74	0.03
	0.7	One	Dichotomous	0.16	0.62	-0.02	0.75	0.19	0.38	0.51	-0.07	0.72	0.45
			Polytomous	0.06	0.58	-0.02	0.75	0.08	0.20	0.39	-0.07	0.72	0.27
		Composite	Dichotomous	0.05	0.72	-0.02	0.75	0.07	0.11	0.67	-0.07	0.72	0.18
			Polytomous	0.05	0.72	-0.02	0.75	0.07	0.11	0.67	-0.07	0.72	0.18
		Separate	Dichotomous	0.05	0.13	-0.02	0.75	0.07	0.11	0.13	-0.07	0.72	0.18
			Polytomous	0.05	0.13	-0.02	0.75	0.07	0.11	0.13	-0.07	0.72	0.18
	1.0	One	Dichotomous	0.15	0.59	0.02	0.75	0.13	0.20	0.78	-0.03	0.73	0.23
			Polytomous	0.06	0.55	0.02	0.75	0.04	0.22	0.51	-0.03	0.73	0.26
		Composite	Dichotomous	0.01	0.72	0.02	0.75	-0.01	0.07	0.70	-0.03	0.73	0.10
			Polytomous	0.01	0.72	0.02	0.75	-0.01	0.07	0.70	-0.03	0.73	0.10
		Separate	Dichotomous	0.01	0.13	0.02	0.75	-0.01	0.07	0.13	-0.03	0.73	0.10
			Polytomous	0.01	0.13	0.02	0.75	-0.01	0.07	0.13	-0.03	0.73	0.10
None	0.2	One	Dichotomous	0.15	0.59	0.01	0.77	0.14	-0.03	1.00	0.02	0.75	-0.04
			Polytomous	0.08	0.55	0.01	0.77	0.07	0.15	0.61	0.02	0.75	0.13
		Composite	Dichotomous	0.02	0.73	0.01	0.77	0.00	0.03	0.72	0.02	0.75	0.01
			Polytomous	0.02	0.73	0.01	0.77	0.00	0.03	0.72	0.02	0.75	0.01
		Separate	Dichotomous	0.02	0.13	0.01	0.77	0.00	0.03	0.13	0.02	0.75	0.01
			Polytomous	0.02	0.13	0.01	0.77	0.00	0.03	0.13	0.02	0.75	0.01
	0.5	One	Dichotomous	0.04	0.69	0.02	0.75	0.02	0.07	0.66	0.01	0.73	0.05
			Polytomous	0.02	0.65	0.02	0.75	0.00	0.03	0.59	0.01	0.73	0.02
		Composite	Dichotomous	0.01	0.72	0.02	0.75	0.00	0.03	0.70	0.01	0.73	0.02
			Polytomous	0.01	0.72	0.02	0.75	0.00	0.03	0.70	0.01	0.73	0.02
		Separate	Dichotomous	0.01	0.13	0.02	0.75	0.00	0.03	0.13	0.01	0.73	0.02
			Polytomous	0.01	0.13	0.02	0.75	0.00	0.03	0.13	0.01	0.73	0.02
	0.7	One	Dichotomous	0.02	0.70	0.03	0.75	-0.01	0.07	0.68	-0.01	0.74	0.08
			Polytomous	0.00	0.67	0.03	0.75	-0.02	0.04	0.64	-0.01	0.74	0.05
		Composite	Dichotomous	0.01	0.71	0.03	0.75	-0.02	0.05	0.70	-0.01	0.74	0.07
			Polytomous	0.01	0.71	0.03	0.75	-0.02	0.05	0.70	-0.01	0.74	0.07
		Separate	Dichotomous	0.01	0.13	0.03	0.75	-0.02	0.05	0.13	-0.01	0.74	0.07

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	0.01	0.13	0.03	0.75	-0.02	0.05	0.13	-0.01	0.74	0.07
	1.0	One	Dichotomous	0.06	0.70	-0.02	0.74	0.07	0.04	0.72	0.01	0.76	0.03
			Polytomous	0.04	0.70	-0.02	0.74	0.06	0.03	0.70	0.01	0.76	0.03
		Composite	Dichotomous	0.05	0.71	-0.02	0.74	0.07	0.03	0.73	0.01	0.76	0.03
			Polytomous	0.05	0.71	-0.02	0.74	0.07	0.03	0.73	0.01	0.76	0.03
		Separate	Dichotomous	0.05	0.13	-0.02	0.74	0.07	0.03	0.13	0.01	0.76	0.03
			Polytomous	0.05	0.13	-0.02	0.74	0.07	0.03	0.13	0.01	0.76	0.03

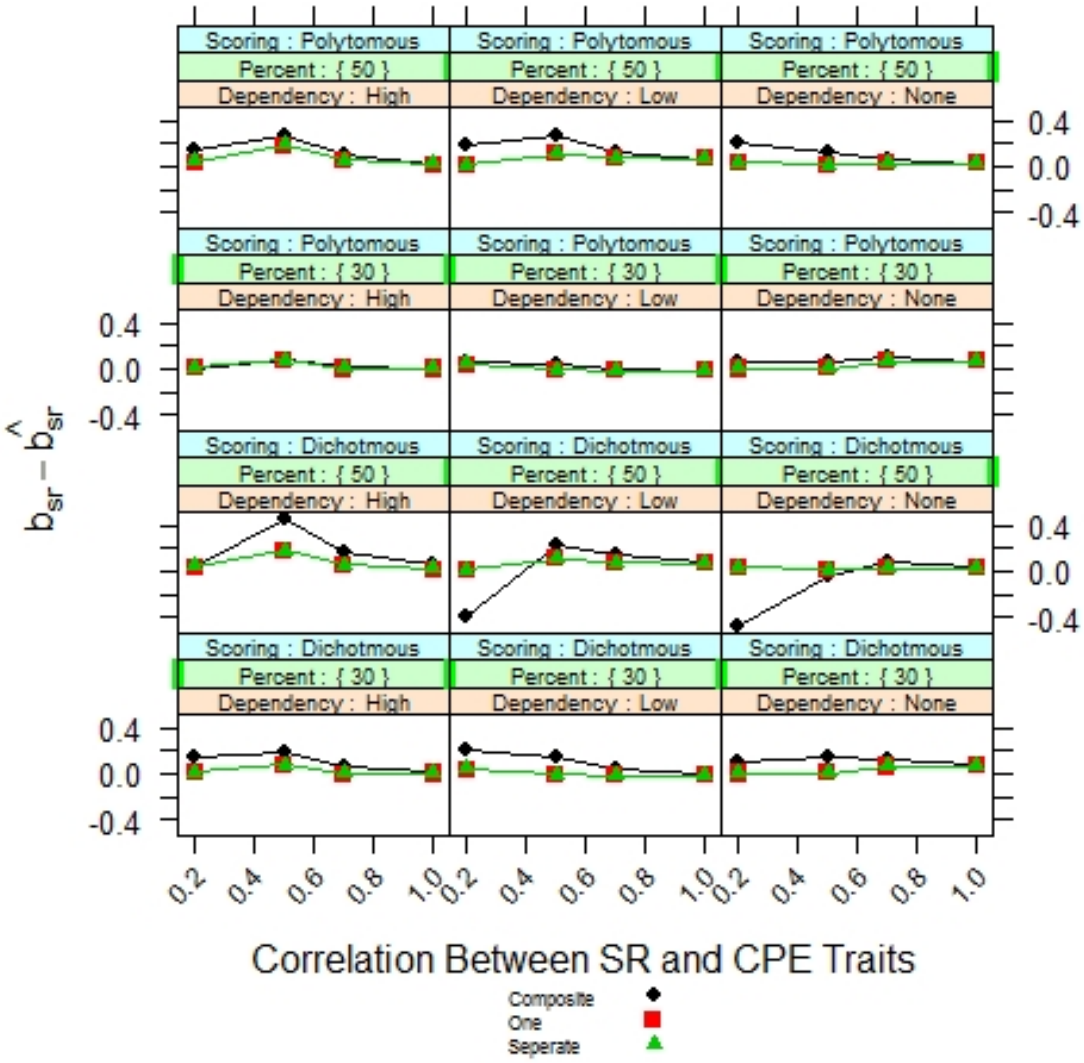

Figure 20: Difference Between Estimated and True SR "b" Parameters for Sample Size of 3000 and 60 Items

Table 21: Average Estimated and True Selected Response *b* Parameter for 3000 Test-Takers and 120 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.03	0.72	-0.01	0.75	0.04	0.02	0.71	0.00	0.74	0.02
			Polytomous	0.03	0.72	-0.01	0.75	0.04	0.02	0.71	0.00	0.74	0.02
		Composite	Dichotomous	0.11	0.66	-0.01	0.75	0.12	0.15	0.64	0.00	0.74	0.16
			Polytomous	0.05	0.59	-0.01	0.75	0.06	0.08	0.50	0.00	0.74	0.08
		Separate	Dichotomous	0.03	0.13	-0.01	0.75	0.04	0.02	0.13	0.00	0.74	0.02
			Polytomous	0.03	0.13	-0.01	0.75	0.04	0.02	0.13	0.00	0.74	0.02
	0.5	One	Dichotomous	0.03	0.72	-0.01	0.75	0.04	0.02	0.72	0.00	0.75	0.02
			Polytomous	0.03	0.72	-0.01	0.75	0.04	0.02	0.72	0.00	0.75	0.02
		Composite	Dichotomous	0.10	0.68	-0.01	0.75	0.11	0.11	0.72	0.00	0.75	0.10
			Polytomous	0.06	0.60	-0.01	0.75	0.07	0.07	0.56	0.00	0.75	0.06
		Separate	Dichotomous	0.03	0.13	-0.01	0.75	0.04	0.02	0.13	0.00	0.75	0.02
			Polytomous	0.03	0.13	-0.01	0.75	0.04	0.02	0.13	0.00	0.75	0.02
	0.7	One	Dichotomous	0.01	0.72	0.01	0.76	0.01	0.04	0.69	-0.02	0.72	0.06
			Polytomous	0.01	0.72	0.01	0.76	0.01	0.04	0.69	-0.02	0.72	0.06
		Composite	Dichotomous	0.07	0.69	0.01	0.76	0.07	0.12	0.72	-0.02	0.72	0.14
			Polytomous	0.04	0.62	0.01	0.76	0.04	0.08	0.60	-0.02	0.72	0.10
		Separate	Dichotomous	0.01	0.13	0.01	0.76	0.01	0.04	0.13	-0.02	0.72	0.06
			Polytomous	0.01	0.13	0.01	0.76	0.01	0.04	0.13	-0.02	0.72	0.06
	1.0	One	Dichotomous	0.00	0.71	0.02	0.74	-0.01	0.02	0.70	0.00	0.74	0.03
			Polytomous	0.00	0.71	0.02	0.74	-0.01	0.02	0.70	0.00	0.74	0.03
		Composite	Dichotomous	0.20	0.42	0.02	0.74	0.19	0.35	0.50	0.00	0.74	0.35
			Polytomous	0.03	0.52	0.02	0.74	0.01	0.14	0.33	0.00	0.74	0.14
		Separate	Dichotomous	0.00	0.13	0.02	0.74	-0.01	0.02	0.13	0.00	0.74	0.03
			Polytomous	0.00	0.13	0.02	0.74	-0.01	0.02	0.13	0.00	0.74	0.03
Low	0.2	One	Dichotomous	0.02	0.73	0.00	0.76	0.02	0.02	0.71	0.00	0.75	0.02
			Polytomous	0.02	0.73	0.00	0.76	0.02	0.02	0.71	0.00	0.75	0.02
		Composite	Dichotomous	0.22	0.46	0.00	0.76	0.22	-0.25	1.60	0.00	0.75	-0.25
			Polytomous	0.07	0.49	0.00	0.76	0.08	0.18	0.35	0.00	0.75	0.18
		Separate	Dichotomous	0.02	0.12	0.00	0.76	0.02	0.02	0.13	0.00	0.75	0.02
			Polytomous	0.02	0.12	0.00	0.76	0.02	0.02	0.13	0.00	0.75	0.02
	0.5	One	Dichotomous	0.04	0.72	-0.03	0.74	0.07	0.02	0.73	0.01	0.76	0.01
			Polytomous	0.04	0.72	-0.03	0.74	0.07	0.02	0.73	0.01	0.76	0.01
		Composite	Dichotomous	0.27	0.47	-0.03	0.74	0.30	-0.41	1.85	0.01	0.76	-0.42
			Polytomous	0.12	0.44	-0.03	0.74	0.15	0.22	0.58	0.01	0.76	0.21
		Separate	Dichotomous	0.04	0.13	-0.03	0.74	0.07	0.02	0.13	0.01	0.76	0.01
			Polytomous	0.04	0.13	-0.03	0.74	0.07	0.02	0.13	0.01	0.76	0.01
	0.7	One	Dichotomous	0.00	0.71	0.02	0.74	-0.02	0.02	0.72	0.00	0.75	0.02
			Polytomous	0.00	0.71	0.02	0.74	-0.02	0.02	0.72	0.00	0.75	0.02
		Composite	Dichotomous	0.14	0.59	0.02	0.74	0.12	0.33	0.66	0.00	0.75	0.33
			Polytomous	0.03	0.54	0.02	0.74	0.01	0.13	0.43	0.00	0.75	0.13
		Separate	Dichotomous	0.00	0.13	0.02	0.74	-0.02	0.02	0.13	0.00	0.75	0.02
			Polytomous	0.00	0.13	0.02	0.74	-0.02	0.02	0.13	0.00	0.75	0.02
	1.0	One	Dichotomous	0.03	0.72	-0.01	0.75	0.04	0.00	0.72	0.02	0.75	-0.02
			Polytomous	0.03	0.72	-0.01	0.75	0.04	0.00	0.72	0.02	0.75	-0.02
		Composite	Dichotomous	0.18	0.61	-0.01	0.75	0.19	0.12	0.85	0.02	0.75	0.09
			Polytomous	0.09	0.54	-0.01	0.75	0.10	0.17	0.51	0.02	0.75	0.15
		Separate	Dichotomous	0.03	0.13	-0.01	0.75	0.04	0.00	0.13	0.02	0.75	-0.02
			Polytomous	0.03	0.13	-0.01	0.75	0.04	0.00	0.13	0.02	0.75	-0.02
None	0.2	One	Dichotomous	0.03	0.71	-0.01	0.75	0.04	0.03	0.74	-0.01	0.76	0.04
			Polytomous	0.03	0.71	-0.01	0.75	0.04	0.03	0.74	-0.01	0.76	0.04
		Composite	Dichotomous	0.19	0.59	-0.01	0.75	0.20	0.04	1.02	-0.01	0.76	0.05
			Polytomous	0.10	0.52	-0.01	0.75	0.12	0.18	0.64	-0.01	0.76	0.18
		Separate	Dichotomous	0.03	0.13	-0.01	0.75	0.04	0.03	0.13	-0.01	0.76	0.04
			Polytomous	0.03	0.13	-0.01	0.75	0.04	0.03	0.13	-0.01	0.76	0.04
	0.5	One	Dichotomous	0.05	0.71	-0.03	0.74	0.08	0.00	0.71	0.03	0.74	-0.04
			Polytomous	0.05	0.71	-0.03	0.74	0.08	0.00	0.71	0.03	0.74	-0.04
		Composite	Dichotomous	0.08	0.68	-0.03	0.74	0.11	0.04	0.68	0.03	0.74	0.01
			Polytomous	0.04	0.64	-0.03	0.74	0.07	0.00	0.60	0.03	0.74	-0.03
		Separate	Dichotomous	0.05	0.13	-0.03	0.74	0.08	0.00	0.13	0.03	0.74	-0.04
			Polytomous	0.05	0.13	-0.03	0.74	0.08	0.00	0.13	0.03	0.74	-0.04
	0.7	One	Dichotomous	0.02	0.70	0.00	0.73	0.02	0.00	0.72	0.02	0.76	-0.02
			Polytomous	0.02	0.70	0.00	0.73	0.02	0.00	0.72	0.02	0.76	-0.02
		Composite	Dichotomous	0.04	0.69	0.00	0.73	0.03	0.03	0.70	0.02	0.76	0.01
			Polytomous	0.01	0.67	0.00	0.73	0.01	0.01	0.66	0.02	0.76	-0.02

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
		Separate	Dichotomous	0.02	0.13	0.00	0.73	0.02	0.00	0.13	0.02	0.76	-0.02
			Polytomous	0.02	0.13	0.00	0.73	0.02	0.00	0.13	0.02	0.76	-0.02
	1.0	One	Dichotomous	0.03	0.71	-0.01	0.74	0.04	0.04	0.72	-0.02	0.75	0.06
			Polytomous	0.03	0.71	-0.01	0.74	0.04	0.04	0.72	-0.02	0.75	0.06
		Composite	Dichotomous	0.04	0.70	-0.01	0.74	0.05	0.06	0.71	-0.02	0.75	0.08
			Polytomous	0.02	0.69	-0.01	0.74	0.03	0.05	0.69	-0.02	0.75	0.06
		Separate	Dichotomous	0.03	0.13	-0.01	0.74	0.04	0.04	0.13	-0.02	0.75	0.06
			Polytomous	0.03	0.13	-0.01	0.74	0.04	0.04	0.13	-0.02	0.75	0.06

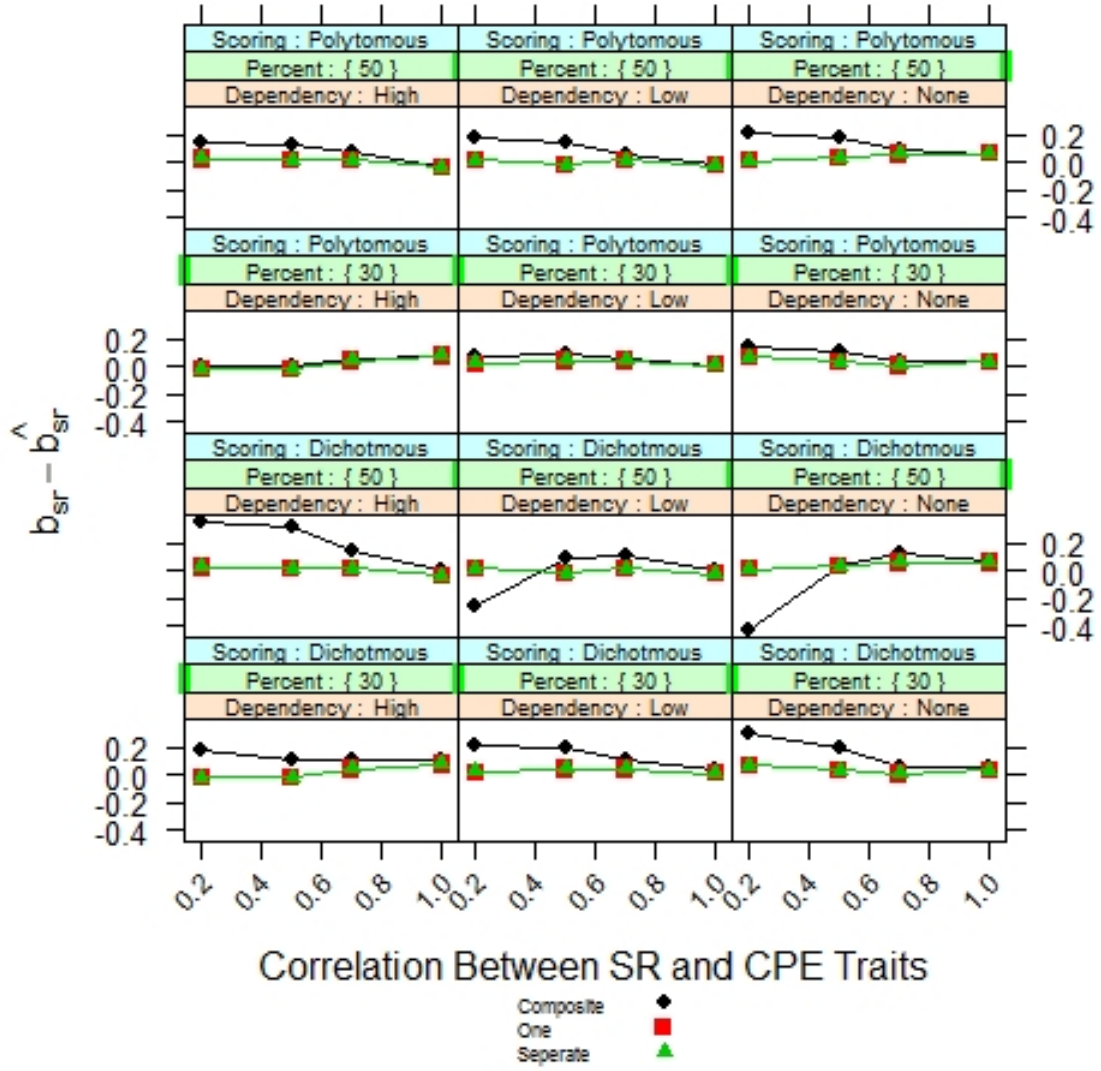

Figure 21: Difference Between Estimated and True SR "b" Parameters for Sample Size of 3000 and 120 Items

Table 22: Average Estimated and True CPE *b* Parameter for 1000 Test- Takers and 60 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.36	0.75	-0.03	0.74	0.39	0.18	0.74	0.05	0.73	0.14
			Polytomous	0.00	0.41	-0.03	0.74	0.04	-0.06	0.41	0.05	0.73	-0.11
		Composite	Dichotomous	0.34	0.75	-0.03	0.74	0.37	0.16	0.75	0.05	0.73	0.11
			Polytomous	-0.01	0.37	-0.03	0.74	0.02	-0.06	0.36	0.05	0.73	-0.10
		Separate	Dichotomous	0.24	0.56	-0.03	0.74	0.28	0.16	0.57	0.05	0.73	0.11
			Polytomous	0.00	0.23	-0.03	0.74	0.04	-0.01	0.23	0.05	0.73	-0.05
	0.5	One	Dichotomous	0.26	0.75	-0.04	0.74	0.30	0.14	0.76	-0.02	0.74	0.15
			Polytomous	0.03	0.44	-0.04	0.74	0.07	0.00	0.40	-0.02	0.74	0.01
		Composite	Dichotomous	0.25	0.76	-0.04	0.74	0.29	0.13	0.77	-0.02	0.74	0.15
			Polytomous	0.00	0.40	-0.04	0.74	0.04	-0.02	0.36	-0.02	0.74	0.00
		Separate	Dichotomous	0.20	0.63	-0.04	0.74	0.24	0.13	0.63	-0.02	0.74	0.15
			Polytomous	0.01	0.29	-0.04	0.74	0.05	0.00	0.27	-0.02	0.74	0.02
	0.7	One	Dichotomous	0.20	0.77	-0.04	0.74	0.24	0.06	0.81	0.02	0.76	0.04
			Polytomous	0.01	0.40	-0.04	0.74	0.06	-0.04	0.42	0.02	0.76	-0.07
		Composite	Dichotomous	0.19	0.78	-0.04	0.74	0.24	0.06	0.83	0.02	0.76	0.03
			Polytomous	0.00	0.37	-0.04	0.74	0.04	-0.05	0.39	0.02	0.76	-0.07
		Separate	Dichotomous	0.15	0.69	-0.04	0.74	0.20	0.06	0.72	0.02	0.76	0.04
			Polytomous	0.00	0.30	-0.04	0.74	0.05	0.00	0.31	0.02	0.76	-0.03
	1.0	One	Dichotomous	1.14	1.91	0.03	0.74	1.11	1.23	1.81	-0.02	0.72	1.25
			Polytomous	-0.11	1.19	0.03	0.74	-0.14	0.02	1.23	-0.02	0.72	0.04
		Composite	Dichotomous	0.82	1.83	0.03	0.74	0.79	0.42	1.61	-0.02	0.72	0.44
			Polytomous	-0.14	1.05	0.03	0.74	-0.17	0.01	1.01	-0.02	0.72	0.03
		Separate	Dichotomous	0.20	0.57	0.03	0.74	0.17	0.19	0.53	-0.02	0.72	0.22
			Polytomous	0.00	0.22	0.03	0.74	-0.03	0.00	0.23	-0.02	0.72	0.02
Low	0.2	One	Dichotomous	0.81	1.73	0.04	0.70	0.77	0.88	1.83	-0.03	0.76	0.91
			Polytomous	-0.14	1.16	0.04	0.70	-0.18	0.07	1.30	-0.03	0.76	0.10
		Composite	Dichotomous	0.28	1.65	0.04	0.70	0.25	0.18	1.76	-0.03	0.76	0.21
			Polytomous	-0.19	1.02	0.04	0.70	-0.23	-0.01	1.07	-0.03	0.76	0.02
		Separate	Dichotomous	0.13	0.60	0.04	0.70	0.09	0.14	0.66	-0.03	0.76	0.17
			Polytomous	-0.01	0.27	0.04	0.70	-0.05	0.00	0.29	-0.03	0.76	0.03
	0.5	One	Dichotomous	0.71	1.80	0.01	0.73	0.70	0.59	1.76	0.01	0.73	0.58
			Polytomous	-0.03	1.13	0.01	0.73	-0.04	-0.07	1.15	0.01	0.73	-0.08
		Composite	Dichotomous	0.29	1.83	0.01	0.73	0.28	0.01	1.75	0.01	0.73	0.00
			Polytomous	-0.12	0.98	0.01	0.73	-0.13	-0.17	0.96	0.01	0.73	-0.18
		Separate	Dichotomous	0.12	0.70	0.01	0.73	0.11	0.08	0.69	0.01	0.73	0.07
			Polytomous	0.00	0.29	0.01	0.73	-0.01	-0.01	0.29	0.01	0.73	-0.02
	0.7	One	Dichotomous	0.47	0.92	0.02	0.72	0.45	0.43	0.98	-0.03	0.76	0.46
			Polytomous	-0.05	0.56	0.02	0.72	-0.07	0.02	0.54	-0.03	0.76	0.05
		Composite	Dichotomous	0.33	0.99	0.02	0.72	0.31	0.21	1.10	-0.03	0.76	0.25
			Polytomous	-0.07	0.51	0.02	0.72	-0.10	-0.02	0.47	-0.03	0.76	0.01
		Separate	Dichotomous	0.21	0.55	0.02	0.72	0.19	0.20	0.59	-0.03	0.76	0.24
			Polytomous	0.00	0.23	0.02	0.72	-0.03	0.00	0.23	-0.03	0.76	0.04
	1.0	One	Dichotomous	0.33	0.97	0.00	0.77	0.33	0.24	0.92	-0.01	0.72	0.25
			Polytomous	-0.02	0.56	0.00	0.77	-0.02	-0.01	0.50	-0.01	0.72	0.00
		Composite	Dichotomous	0.22	1.07	0.00	0.77	0.22	0.06	1.06	-0.01	0.72	0.07
			Polytomous	-0.07	0.52	0.00	0.77	-0.07	-0.06	0.47	-0.01	0.72	-0.05
		Separate	Dichotomous	0.15	0.66	0.00	0.77	0.16	0.13	0.61	-0.01	0.72	0.14
			Polytomous	0.00	0.28	0.00	0.77	0.00	0.00	0.25	-0.01	0.72	0.01
None	0.2	One	Dichotomous	0.26	0.92	0.01	0.72	0.25	0.18	0.88	-0.02	0.70	0.20
			Polytomous	-0.03	0.51	0.01	0.72	-0.04	-0.01	0.48	-0.02	0.70	0.01
		Composite	Dichotomous	0.16	1.01	0.01	0.72	0.15	0.05	1.01	-0.02	0.70	0.06
			Polytomous	-0.08	0.49	0.01	0.72	-0.09	-0.07	0.46	-0.02	0.70	-0.05
		Separate	Dichotomous	0.13	0.69	0.01	0.72	0.12	0.10	0.66	-0.02	0.70	0.11
			Polytomous	0.00	0.28	0.01	0.72	-0.01	0.00	0.28	-0.02	0.70	0.02
	0.5	One	Dichotomous	0.20	0.60	0.00	0.74	0.21	0.10	0.59	0.03	0.74	0.07
			Polytomous	-0.02	0.32	0.00	0.74	-0.01	-0.04	0.30	0.03	0.74	-0.07
		Composite	Dichotomous	0.21	0.59	0.00	0.74	0.22	0.12	0.57	0.03	0.74	0.09
			Polytomous	-0.02	0.30	0.00	0.74	-0.01	-0.03	0.27	0.03	0.74	-0.06
		Separate	Dichotomous	0.22	0.58	0.00	0.74	0.22	0.17	0.55	0.03	0.74	0.13
			Polytomous	0.00	0.24	0.00	0.74	0.00	0.00	0.23	0.03	0.74	-0.04
	0.7	One	Dichotomous	0.10	0.63	0.03	0.72	0.07	0.11	0.67	-0.03	0.77	0.14
			Polytomous	-0.04	0.32	0.03	0.72	-0.07	0.01	0.30	-0.03	0.77	0.04
		Composite	Dichotomous	0.10	0.63	0.03	0.72	0.07	0.11	0.66	-0.03	0.77	0.14
			Polytomous	-0.03	0.30	0.03	0.72	-0.06	0.01	0.28	-0.03	0.77	0.04
		Separate	Dichotomous	0.14	0.61	0.03	0.72	0.11	0.14	0.65	-0.03	0.77	0.17

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	0.00	0.27	0.03	0.72	-0.03	0.01	0.26	-0.03	0.77	0.04
	1.0	One	Dichotomous	0.11	0.73	-0.02	0.79	0.13	0.06	0.67	0.00	0.72	0.06
			Polytomous	0.00	0.35	-0.02	0.79	0.01	-0.01	0.30	0.00	0.72	-0.01
		Composite	Dichotomous	0.10	0.73	-0.02	0.79	0.12	0.07	0.67	0.00	0.72	0.07
			Polytomous	0.00	0.34	-0.02	0.79	0.01	-0.01	0.29	0.00	0.72	0.00
		Separate	Dichotomous	0.14	0.73	-0.02	0.79	0.15	0.09	0.68	0.00	0.72	0.09
			Polytomous	0.00	0.32	-0.02	0.79	0.02	0.00	0.28	0.00	0.72	0.00

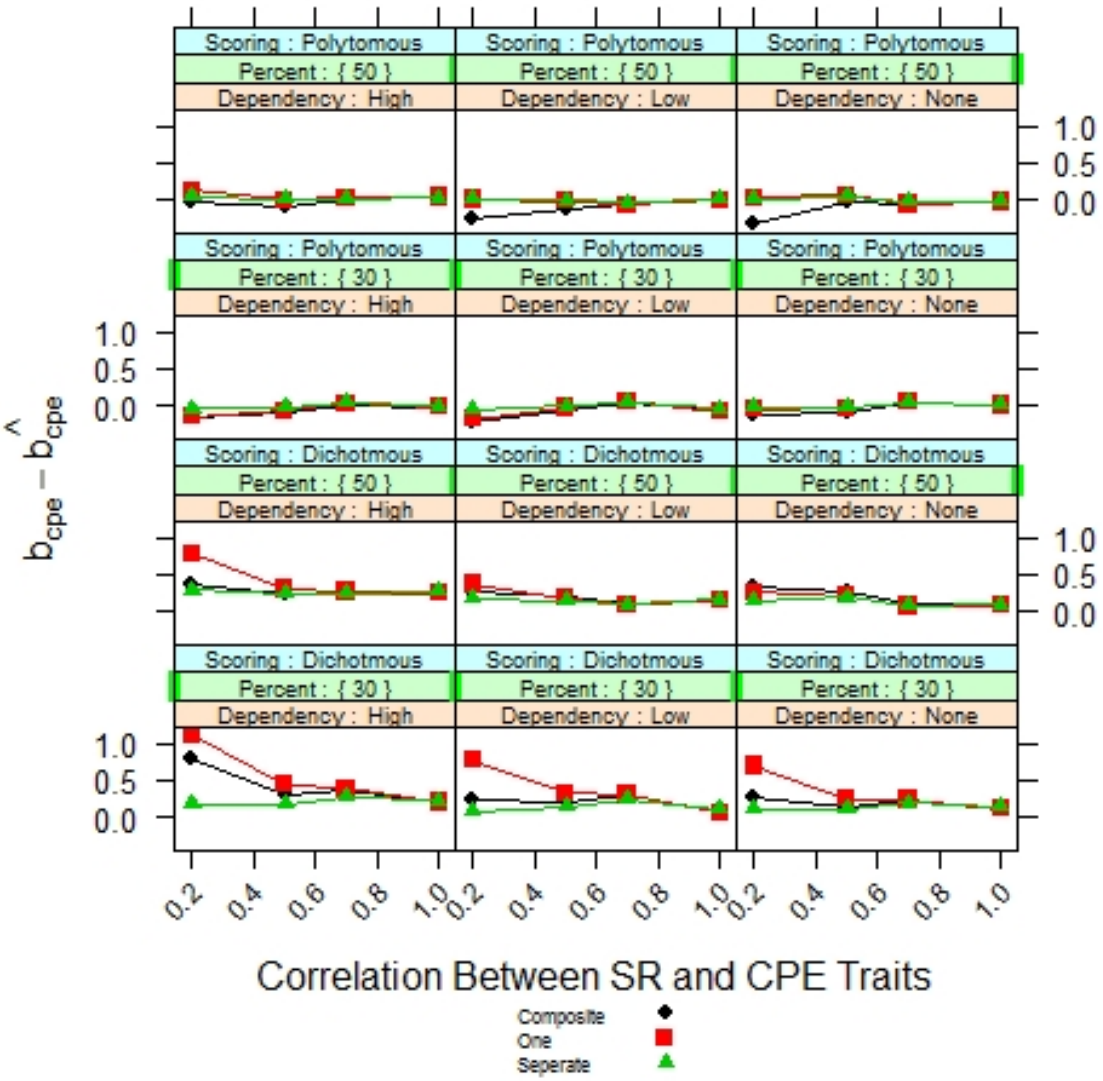

Figure 22: Difference Between Estimated and True CPE "b" Parameters for Sample Size of 1000 and 60 Items

Table 23: Average Estimated and True CPE *b* Parameter for 1000 Test- Takers and 120 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.30	0.76	0.01	0.75	0.29	0.21	0.76	0.00	0.74	0.21
			Polytomous	-0.04	0.44	0.01	0.75	-0.05	-0.03	0.41	0.00	0.74	-0.02
		Composite	Dichotomous	0.30	0.78	0.01	0.75	0.30	0.22	0.77	0.00	0.74	0.22
			Polytomous	-0.05	0.39	0.01	0.75	-0.05	-0.04	0.36	0.00	0.74	-0.03
		Separate	Dichotomous	0.22	0.53	0.01	0.75	0.21	0.18	0.51	0.00	0.74	0.18
			Polytomous	0.00	0.23	0.01	0.75	0.00	0.00	0.22	0.00	0.74	0.00
	0.5	One	Dichotomous	0.19	0.76	-0.01	0.73	0.20	0.12	0.78	-0.01	0.75	0.12
			Polytomous	-0.03	0.40	-0.01	0.73	-0.02	-0.02	0.40	-0.01	0.75	-0.02
		Composite	Dichotomous	0.20	0.78	-0.01	0.73	0.21	0.13	0.79	-0.01	0.75	0.14
			Polytomous	-0.03	0.36	-0.01	0.73	-0.02	-0.03	0.36	-0.01	0.75	-0.03
		Separate	Dichotomous	0.16	0.62	-0.01	0.73	0.17	0.13	0.62	-0.01	0.75	0.13
			Polytomous	0.00	0.26	-0.01	0.73	0.01	0.00	0.26	-0.01	0.75	0.00
	0.7	One	Dichotomous	0.16	0.82	-0.03	0.76	0.19	0.08	0.81	-0.01	0.76	0.09
			Polytomous	-0.01	0.40	-0.03	0.76	0.02	-0.02	0.39	-0.01	0.76	-0.01
		Composite	Dichotomous	0.17	0.85	-0.03	0.76	0.20	0.10	0.82	-0.01	0.76	0.11
			Polytomous	-0.02	0.37	-0.03	0.76	0.01	-0.03	0.36	-0.01	0.76	-0.02
		Separate	Dichotomous	0.14	0.72	-0.03	0.76	0.17	0.09	0.69	-0.01	0.76	0.10
			Polytomous	0.00	0.29	-0.03	0.76	0.03	0.00	0.28	-0.01	0.76	0.01
	1.0	One	Dichotomous	1.15	2.00	0.01	0.76	1.14	1.12	1.98	0.02	0.76	1.11
			Polytomous	-0.06	1.26	0.01	0.76	-0.07	-0.08	1.31	0.02	0.76	-0.10
		Composite	Dichotomous	0.74	1.88	0.01	0.76	0.73	0.31	1.69	0.02	0.76	0.29
			Polytomous	-0.11	1.06	0.01	0.76	-0.12	-0.12	1.10	0.02	0.76	-0.14
		Separate	Dichotomous	0.23	0.54	0.01	0.76	0.21	0.17	0.52	0.02	0.76	0.15
			Polytomous	0.00	0.23	0.01	0.76	-0.01	0.00	0.24	0.02	0.76	-0.02
Low	0.2	One	Dichotomous	0.76	1.80	-0.01	0.75	0.78	0.76	1.78	-0.01	0.73	0.77
			Polytomous	-0.05	1.07	-0.01	0.75	-0.04	-0.01	1.19	-0.01	0.73	0.01
		Composite	Dichotomous	0.37	1.78	-0.01	0.75	0.38	0.11	1.64	-0.01	0.73	0.12
			Polytomous	-0.12	0.90	-0.01	0.75	-0.11	-0.12	0.97	-0.01	0.73	-0.10
		Separate	Dichotomous	0.15	0.63	-0.01	0.75	0.16	0.12	0.61	-0.01	0.73	0.14
			Polytomous	0.00	0.24	-0.01	0.75	0.01	0.00	0.27	-0.01	0.73	0.01
	0.5	One	Dichotomous	0.73	1.76	-0.02	0.74	0.75	0.60	1.84	-0.01	0.75	0.61
			Polytomous	-0.01	1.14	-0.02	0.74	0.01	-0.02	1.10	-0.01	0.75	-0.02
		Composite	Dichotomous	0.26	1.81	-0.02	0.74	0.28	-0.07	1.79	-0.01	0.75	-0.06
			Polytomous	-0.12	0.95	-0.02	0.74	-0.10	-0.16	0.89	-0.01	0.75	-0.16
		Separate	Dichotomous	0.13	0.68	-0.02	0.74	0.15	0.09	0.70	-0.01	0.75	0.09
			Polytomous	0.00	0.28	-0.02	0.74	0.02	0.00	0.28	-0.01	0.75	0.01
	0.7	One	Dichotomous	0.51	1.02	-0.04	0.77	0.56	0.40	0.96	-0.03	0.73	0.42
			Polytomous	0.00	0.61	-0.04	0.77	0.04	0.00	0.54	-0.03	0.73	0.03
		Composite	Dichotomous	0.42	1.11	-0.04	0.77	0.46	0.24	1.07	-0.03	0.73	0.27
			Polytomous	-0.05	0.56	-0.04	0.77	-0.01	-0.04	0.49	-0.03	0.73	-0.02
		Separate	Dichotomous	0.25	0.54	-0.04	0.77	0.30	0.20	0.51	-0.03	0.73	0.23
			Polytomous	0.00	0.23	-0.04	0.77	0.04	0.00	0.21	-0.03	0.73	0.03
	1.0	One	Dichotomous	0.28	0.96	0.01	0.74	0.27	0.17	0.99	0.01	0.75	0.15
			Polytomous	-0.06	0.54	0.01	0.74	-0.07	-0.05	0.52	0.01	0.75	-0.07
		Composite	Dichotomous	0.18	1.08	0.01	0.74	0.16	0.01	1.17	0.01	0.75	-0.01
			Polytomous	-0.11	0.52	0.01	0.74	-0.12	-0.10	0.50	0.01	0.75	-0.12
		Separate	Dichotomous	0.14	0.62	0.01	0.74	0.13	0.10	0.63	0.01	0.75	0.09
			Polytomous	0.00	0.26	0.01	0.74	-0.01	0.00	0.26	0.01	0.75	-0.02
None	0.2	One	Dichotomous	0.20	0.98	0.00	0.76	0.20	0.10	0.99	0.01	0.76	0.09
			Polytomous	-0.03	0.54	0.00	0.76	-0.03	-0.05	0.51	0.01	0.76	-0.06
		Composite	Dichotomous	0.08	1.14	0.00	0.76	0.08	-0.05	1.20	0.01	0.76	-0.06
			Polytomous	-0.11	0.56	0.00	0.76	-0.11	-0.12	0.52	0.01	0.76	-0.13
		Separate	Dichotomous	0.12	0.69	0.00	0.76	0.12	0.07	0.71	0.01	0.76	0.06
			Polytomous	0.00	0.30	0.00	0.76	0.00	0.00	0.29	0.01	0.76	-0.01
	0.5	One	Dichotomous	0.18	0.62	0.00	0.76	0.18	0.15	0.63	-0.04	0.77	0.19
			Polytomous	-0.04	0.32	0.00	0.76	-0.04	0.00	0.33	-0.04	0.77	0.04
		Composite	Dichotomous	0.20	0.61	0.00	0.76	0.20	0.18	0.61	-0.04	0.77	0.22
			Polytomous	-0.03	0.29	0.00	0.76	-0.03	0.00	0.30	-0.04	0.77	0.04
		Separate	Dichotomous	0.22	0.54	0.00	0.76	0.22	0.20	0.53	-0.04	0.77	0.24
			Polytomous	0.00	0.22	0.00	0.76	0.00	0.00	0.23	-0.04	0.77	0.04
	0.7	One	Dichotomous	0.12	0.67	-0.02	0.75	0.14	0.08	0.67	-0.02	0.75	0.09
			Polytomous	-0.02	0.34	-0.02	0.75	-0.01	-0.02	0.32	-0.02	0.75	0.00
		Composite	Dichotomous	0.13	0.66	-0.02	0.75	0.15	0.09	0.65	-0.02	0.75	0.11
			Polytomous	-0.02	0.32	-0.02	0.75	0.00	-0.01	0.30	-0.02	0.75	0.00
		Separate	Dichotomous	0.16	0.63	-0.02	0.75	0.18	0.12	0.62	-0.02	0.75	0.14

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	0.00	0.28	-0.02	0.75	0.02	0.00	0.26	-0.02	0.75	0.02
	1.0	One	Dichotomous	0.05	0.71	0.03	0.75	0.02	0.03	0.71	0.00	0.75	0.03
			Polytomous	-0.06	0.33	0.03	0.75	-0.08	-0.03	0.34	0.00	0.75	-0.03
		Composite	Dichotomous	0.05	0.71	0.03	0.75	0.02	0.05	0.70	0.00	0.75	0.04
			Polytomous	-0.04	0.31	0.03	0.75	-0.07	-0.02	0.32	0.00	0.75	-0.03
		Separate	Dichotomous	0.09	0.70	0.03	0.75	0.06	0.08	0.68	0.00	0.75	0.08
			Polytomous	0.00	0.28	0.03	0.75	-0.03	0.00	0.30	0.00	0.75	0.00

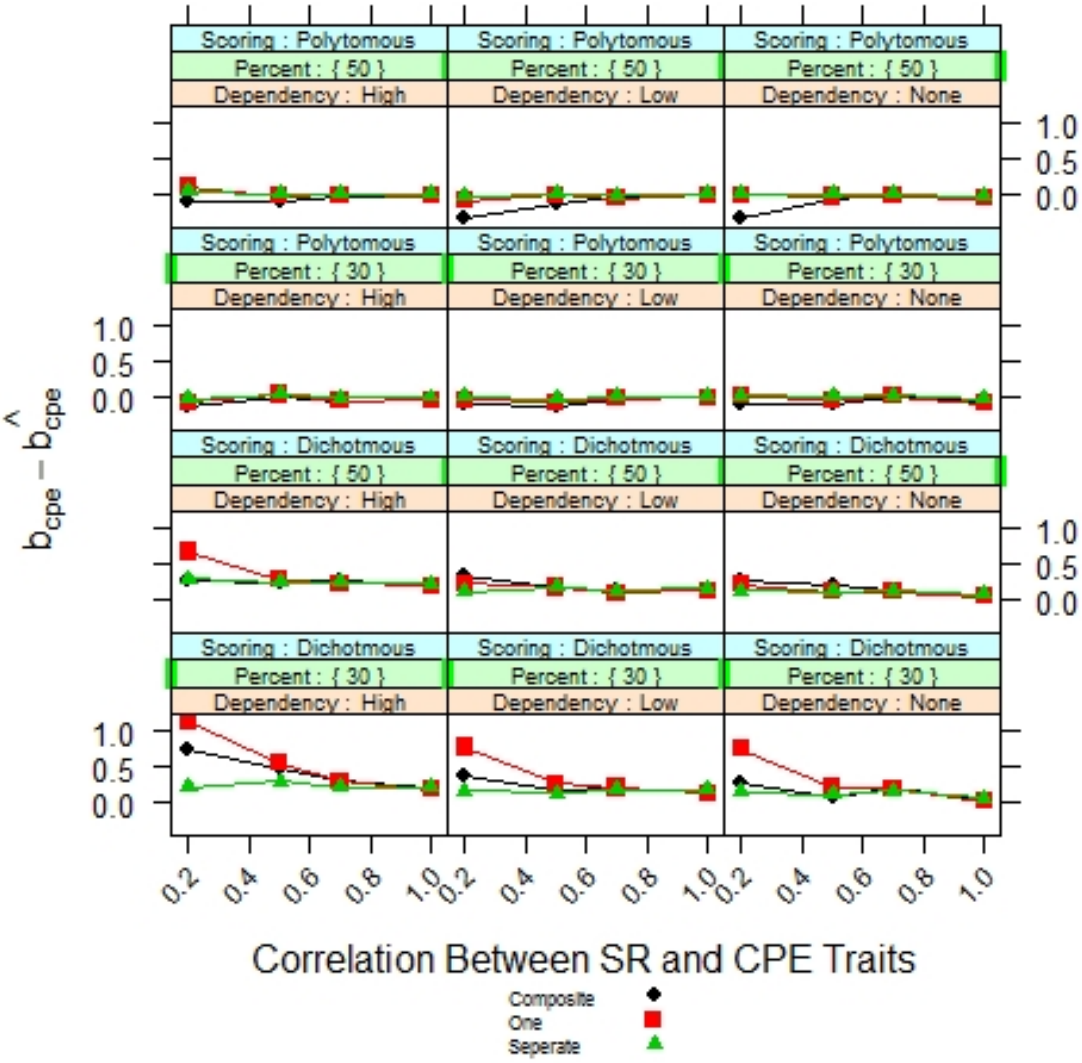

Figure 23: Difference Between Estimated and True CPE "b" Parameters for Sample Size of 1000 and 120 Items

Table 24: Average Estimated and True CPE *b* Parameter for 3000 Test- Takers and 60 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.18	0.74	0.05	0.73	0.14	0.25	0.60	-0.02	0.75	0.27
			Polytomous	-0.06	0.41	0.05	0.73	-0.11	0.01	0.38	-0.02	0.75	0.03
		Composite	Dichotomous	0.16	0.75	0.05	0.73	0.11	0.26	0.62	-0.02	0.75	0.28
			Polytomous	-0.06	0.36	0.05	0.73	-0.10	-0.02	0.32	-0.02	0.75	0.00
		Separate	Dichotomous	0.16	0.57	0.05	0.73	0.11	0.23	0.53	-0.02	0.75	0.25
			Polytomous	-0.01	0.23	0.05	0.73	-0.05	0.00	0.23	-0.02	0.75	0.02
	0.5	One	Dichotomous	0.14	0.76	-0.02	0.74	0.15	0.12	0.62	0.02	0.74	0.10
			Polytomous	0.00	0.40	-0.02	0.74	0.01	-0.04	0.36	0.02	0.74	-0.06
		Composite	Dichotomous	0.13	0.77	-0.02	0.74	0.15	0.14	0.65	0.02	0.74	0.12
			Polytomous	-0.02	0.36	-0.02	0.74	0.00	-0.03	0.31	0.02	0.74	-0.05
		Separate	Dichotomous	0.13	0.63	-0.02	0.74	0.15	0.12	0.63	0.02	0.74	0.10
			Polytomous	0.00	0.27	-0.02	0.74	0.02	-0.01	0.28	0.02	0.74	-0.03
	0.7	One	Dichotomous	0.06	0.81	0.02	0.76	0.04	0.10	0.60	0.02	0.71	0.08
			Polytomous	-0.04	0.42	0.02	0.76	-0.07	-0.04	0.30	0.02	0.71	-0.05
		Composite	Dichotomous	0.06	0.83	0.02	0.76	0.03	0.12	0.64	0.02	0.71	0.10
			Polytomous	-0.05	0.39	0.02	0.76	-0.07	-0.03	0.28	0.02	0.71	-0.05
		Separate	Dichotomous	0.06	0.72	0.02	0.76	0.04	0.10	0.65	0.02	0.71	0.08
			Polytomous	0.00	0.31	0.02	0.76	-0.03	0.00	0.27	0.02	0.71	-0.02
	1.0	One	Dichotomous	1.23	1.81	-0.02	0.72	1.25	0.74	1.10	-0.03	0.73	0.77
			Polytomous	0.02	1.23	-0.02	0.72	0.04	0.09	1.07	-0.03	0.73	0.13
		Composite	Dichotomous	0.42	1.61	-0.02	0.72	0.44	0.34	1.14	-0.03	0.73	0.38
			Polytomous	0.01	1.01	-0.02	0.72	0.03	-0.07	0.80	-0.03	0.73	-0.03
		Separate	Dichotomous	0.19	0.53	-0.02	0.72	0.22	0.26	0.53	-0.03	0.73	0.29
			Polytomous	0.00	0.23	-0.02	0.72	0.02	0.01	0.23	-0.03	0.73	0.04
Low	0.2	One	Dichotomous	0.88	1.83	-0.03	0.76	0.91	0.36	0.82	-0.01	0.75	0.37
			Polytomous	0.07	1.30	-0.03	0.76	0.10	0.00	0.87	-0.01	0.75	0.01
		Composite	Dichotomous	0.18	1.76	-0.03	0.76	0.21	0.27	0.73	-0.01	0.75	0.28
			Polytomous	-0.01	1.07	-0.03	0.76	0.02	-0.25	0.76	-0.01	0.75	-0.24
		Separate	Dichotomous	0.14	0.66	-0.03	0.76	0.17	0.16	0.64	-0.01	0.75	0.17
			Polytomous	0.00	0.29	-0.03	0.76	0.03	0.00	0.27	-0.01	0.75	0.01
	0.5	One	Dichotomous	0.59	1.76	0.01	0.73	0.58	0.24	0.54	-0.02	0.74	0.26
			Polytomous	-0.07	1.15	0.01	0.73	-0.08	0.02	0.70	-0.02	0.74	0.03
		Composite	Dichotomous	0.01	1.75	0.01	0.73	0.00	0.31	0.56	-0.02	0.74	0.32
			Polytomous	-0.17	0.96	0.01	0.73	-0.18	-0.35	0.72	-0.02	0.74	-0.33
		Separate	Dichotomous	0.08	0.69	0.01	0.73	0.07	0.13	0.69	-0.02	0.74	0.15
			Polytomous	-0.01	0.29	0.01	0.73	-0.02	0.00	0.29	-0.02	0.74	0.02
	0.7	One	Dichotomous	0.43	0.98	-0.03	0.76	0.46	0.29	0.65	-0.01	0.73	0.30
			Polytomous	0.02	0.54	-0.03	0.76	0.05	0.00	0.50	-0.01	0.73	0.01
		Composite	Dichotomous	0.21	1.10	-0.03	0.76	0.25	0.22	0.75	-0.01	0.73	0.23
			Polytomous	-0.02	0.47	-0.03	0.76	0.01	-0.11	0.49	-0.01	0.73	-0.09
		Separate	Dichotomous	0.20	0.59	-0.03	0.76	0.24	0.23	0.53	-0.01	0.73	0.24
			Polytomous	0.00	0.23	-0.03	0.76	0.04	0.00	0.24	-0.01	0.73	0.01
	1.0	One	Dichotomous	0.24	0.92	-0.01	0.72	0.25	0.18	0.65	0.00	0.76	0.18
			Polytomous	-0.01	0.50	-0.01	0.72	0.00	-0.02	0.43	0.00	0.76	-0.01
		Composite	Dichotomous	0.06	1.06	-0.01	0.72	0.07	0.20	0.74	0.00	0.76	0.20
			Polytomous	-0.06	0.47	-0.01	0.72	-0.05	-0.13	0.45	0.00	0.76	-0.12
		Separate	Dichotomous	0.13	0.61	-0.01	0.72	0.14	0.14	0.66	0.00	0.76	0.15
			Polytomous	0.00	0.25	-0.01	0.72	0.01	0.00	0.27	0.00	0.76	0.00
None	0.2	One	Dichotomous	0.18	0.88	-0.02	0.70	0.20	0.18	0.62	-0.04	0.74	0.22
			Polytomous	-0.01	0.48	-0.02	0.70	0.01	0.02	0.37	-0.04	0.74	0.06
		Composite	Dichotomous	0.05	1.01	-0.02	0.70	0.06	0.23	0.69	-0.04	0.74	0.28
			Polytomous	-0.07	0.46	-0.02	0.70	-0.05	-0.08	0.39	-0.04	0.74	-0.04
		Separate	Dichotomous	0.10	0.66	-0.02	0.70	0.11	0.15	0.69	-0.04	0.74	0.20
			Polytomous	0.00	0.28	-0.02	0.70	0.02	0.00	0.29	-0.04	0.74	0.05
	0.5	One	Dichotomous	0.10	0.59	0.03	0.74	0.07	0.22	0.56	-0.03	0.73	0.25
			Polytomous	-0.04	0.30	0.03	0.74	-0.07	0.01	0.29	-0.03	0.73	0.04
		Composite	Dichotomous	0.12	0.57	0.03	0.74	0.09	0.23	0.55	-0.03	0.73	0.26
			Polytomous	-0.03	0.27	0.03	0.74	-0.06	0.00	0.25	-0.03	0.73	0.03
		Separate	Dichotomous	0.17	0.55	0.03	0.74	0.13	0.25	0.51	-0.03	0.73	0.28
			Polytomous	0.00	0.23	0.03	0.74	-0.04	0.00	0.22	-0.03	0.73	0.03
	0.7	One	Dichotomous	0.11	0.67	-0.03	0.77	0.14	0.14	0.64	-0.01	0.74	0.14
			Polytomous	0.01	0.30	-0.03	0.77	0.04	0.00	0.32	-0.01	0.74	0.00
		Composite	Dichotomous	0.11	0.66	-0.03	0.77	0.14	0.13	0.64	-0.01	0.74	0.14
			Polytomous	0.01	0.28	-0.03	0.77	0.04	0.00	0.29	-0.01	0.74	0.00
		Separate	Dichotomous	0.14	0.65	-0.03	0.77	0.17	0.16	0.62	-0.01	0.74	0.17

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
			Polytomous	0.01	0.26	-0.03	0.77	0.04	0.00	0.27	-0.01	0.74	0.01
	1.0	One	Dichotomous	0.06	0.67	0.00	0.72	0.06	0.10	0.69	0.01	0.74	0.09
			Polytomous	-0.01	0.30	0.00	0.72	-0.01	-0.01	0.32	0.01	0.74	-0.02
		Composite	Dichotomous	0.07	0.67	0.00	0.72	0.07	0.08	0.68	0.01	0.74	0.08
			Polytomous	-0.01	0.29	0.00	0.72	0.00	-0.01	0.30	0.01	0.74	-0.01
		Separate	Dichotomous	0.09	0.68	0.00	0.72	0.09	0.11	0.68	0.01	0.74	0.10
			Polytomous	0.00	0.28	0.00	0.72	0.00	0.00	0.29	0.01	0.74	-0.01

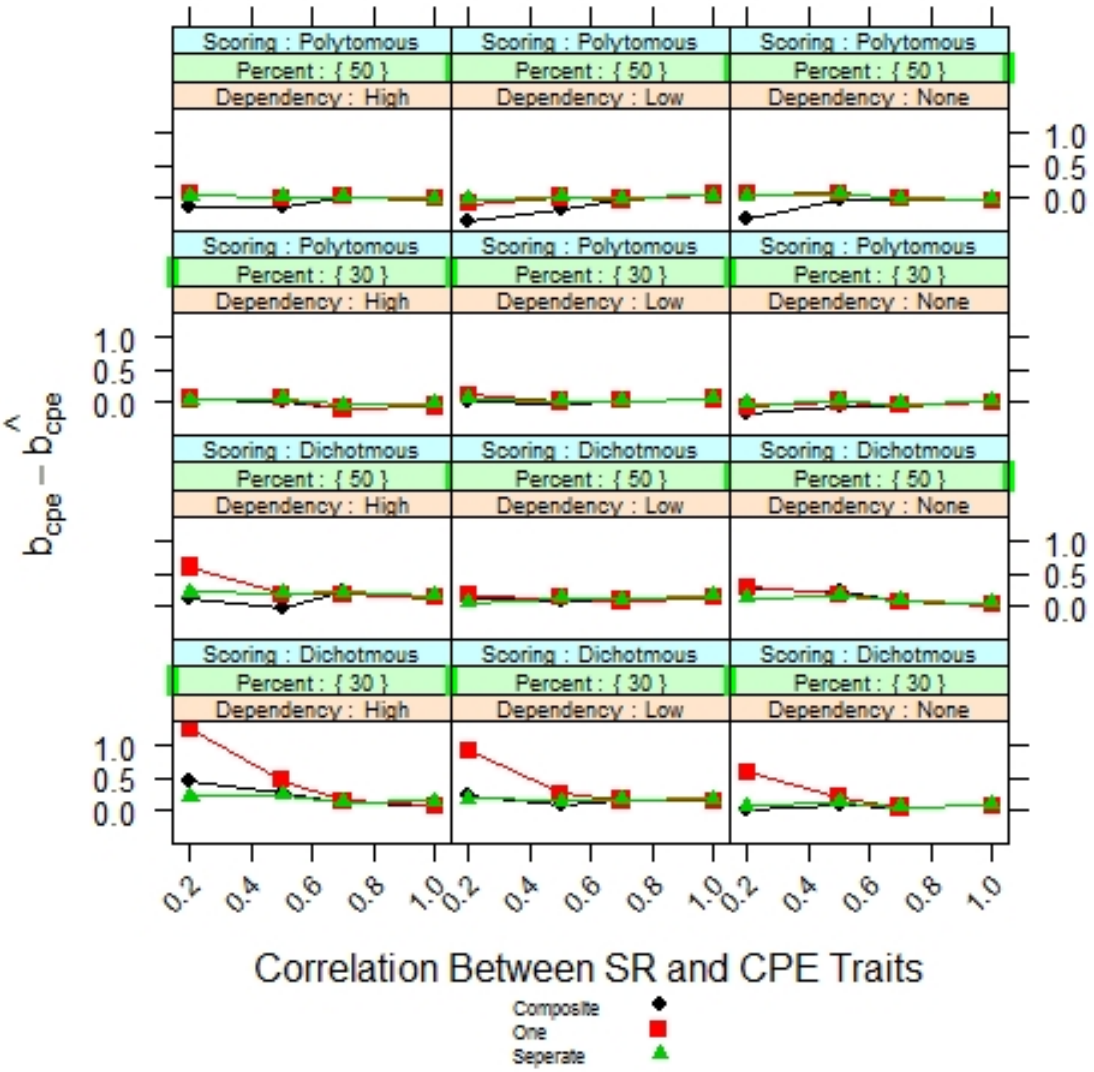

Figure 24: Difference Between Estimated and True CPE "b" Parameters for Sample Size of 3000 and 60 Items

Table 25: Average Estimated and True CPE *b* Parameter for 3000 Test- Takers and 120 Items

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
High	0.2	One	Dichotomous	0.21	0.76	0.00	0.74	0.21	0.22	0.62	-0.01	0.74	0.23
			Polytomous	-0.03	0.41	0.00	0.74	-0.02	-0.02	0.37	-0.01	0.74	-0.01
		Composite	Dichotomous	0.22	0.77	0.00	0.74	0.22	0.26	0.65	-0.01	0.74	0.27
			Polytomous	-0.04	0.36	0.00	0.74	-0.03	-0.03	0.31	-0.01	0.74	-0.01
		Separate	Dichotomous	0.18	0.51	0.00	0.74	0.18	0.24	0.52	-0.01	0.74	0.25
			Polytomous	0.00	0.22	0.00	0.74	0.00	0.00	0.21	-0.01	0.74	0.01
	0.5	One	Dichotomous	0.12	0.78	-0.01	0.75	0.12	0.11	0.63	0.01	0.76	0.10
			Polytomous	-0.02	0.40	-0.01	0.75	-0.02	-0.04	0.35	0.01	0.76	-0.06
		Composite	Dichotomous	0.13	0.79	-0.01	0.75	0.14	0.15	0.68	0.01	0.76	0.13
			Polytomous	-0.03	0.36	-0.01	0.75	-0.03	-0.03	0.30	0.01	0.76	-0.04
		Separate	Dichotomous	0.13	0.62	-0.01	0.75	0.13	0.14	0.63	0.01	0.76	0.12
			Polytomous	0.00	0.26	-0.01	0.75	0.00	0.00	0.26	0.01	0.76	-0.01
	0.7	One	Dichotomous	0.08	0.81	-0.01	0.76	0.09	0.10	0.64	-0.01	0.75	0.11
			Polytomous	-0.02	0.39	-0.01	0.76	-0.01	-0.02	0.33	-0.01	0.75	-0.01
		Composite	Dichotomous	0.10	0.82	-0.01	0.76	0.11	0.14	0.71	-0.01	0.75	0.14
			Polytomous	-0.03	0.36	-0.01	0.76	-0.02	-0.02	0.30	-0.01	0.75	-0.01
		Separate	Dichotomous	0.09	0.69	-0.01	0.76	0.10	0.12	0.69	-0.01	0.75	0.13
			Polytomous	0.00	0.28	-0.01	0.76	0.01	0.00	0.29	-0.01	0.75	0.01
	1.0	One	Dichotomous	1.12	1.98	0.02	0.76	1.11	0.64	1.10	-0.04	0.76	0.68
			Polytomous	-0.08	1.31	0.02	0.76	-0.10	0.07	0.98	-0.04	0.76	0.10
		Composite	Dichotomous	0.31	1.69	0.02	0.76	0.29	0.22	1.28	-0.04	0.76	0.26
			Polytomous	-0.12	1.10	0.02	0.76	-0.14	-0.12	0.77	-0.04	0.76	-0.08
		Separate	Dichotomous	0.17	0.52	0.02	0.76	0.15	0.26	0.53	-0.04	0.76	0.30
			Polytomous	0.00	0.24	0.02	0.76	-0.02	0.00	0.22	-0.04	0.76	0.04
Low	0.2	One	Dichotomous	0.76	1.78	-0.01	0.73	0.77	0.25	0.78	0.02	0.76	0.24
			Polytomous	-0.01	1.19	-0.01	0.73	0.01	-0.07	0.85	0.02	0.76	-0.09
		Composite	Dichotomous	0.11	1.64	-0.01	0.73	0.12	0.36	0.64	0.02	0.76	0.34
			Polytomous	-0.12	0.97	-0.01	0.73	-0.10	-0.32	0.79	0.02	0.76	-0.34
		Separate	Dichotomous	0.12	0.61	-0.01	0.73	0.14	0.14	0.65	0.02	0.76	0.12
			Polytomous	0.00	0.27	-0.01	0.73	0.01	0.00	0.27	0.02	0.76	-0.02
	0.5	One	Dichotomous	0.60	1.84	-0.01	0.75	0.61	0.20	0.51	-0.01	0.73	0.21
			Polytomous	-0.02	1.10	-0.01	0.75	-0.02	-0.02	0.68	-0.01	0.73	-0.01
		Composite	Dichotomous	-0.07	1.79	-0.01	0.75	-0.06	0.27	0.69	-0.01	0.73	0.28
			Polytomous	-0.16	0.89	-0.01	0.75	-0.16	-0.36	0.70	-0.01	0.73	-0.35
		Separate	Dichotomous	0.09	0.70	-0.01	0.75	0.09	0.12	0.70	-0.01	0.73	0.13
			Polytomous	0.00	0.28	-0.01	0.75	0.01	0.00	0.28	-0.01	0.73	0.01
	0.7	One	Dichotomous	0.40	0.96	-0.03	0.73	0.42	0.28	0.65	-0.01	0.74	0.29
			Polytomous	0.00	0.54	-0.03	0.73	0.03	-0.03	0.46	-0.01	0.74	-0.02
		Composite	Dichotomous	0.24	1.07	-0.03	0.73	0.27	0.24	0.78	-0.01	0.74	0.25
			Polytomous	-0.04	0.49	-0.03	0.73	-0.02	-0.10	0.44	-0.01	0.74	-0.09
		Separate	Dichotomous	0.20	0.51	-0.03	0.73	0.23	0.23	0.51	-0.01	0.74	0.24
			Polytomous	0.00	0.21	-0.03	0.73	0.03	0.00	0.21	-0.01	0.74	0.01
	1.0	One	Dichotomous	0.17	0.99	0.01	0.75	0.15	0.16	0.62	-0.02	0.74	0.17
			Polytomous	-0.05	0.52	0.01	0.75	-0.07	-0.02	0.42	-0.02	0.74	0.00
		Composite	Dichotomous	0.01	1.17	0.01	0.75	-0.01	0.17	0.76	-0.02	0.74	0.18
			Polytomous	-0.10	0.50	0.01	0.75	-0.12	-0.14	0.46	-0.02	0.74	-0.13
		Separate	Dichotomous	0.10	0.63	0.01	0.75	0.09	0.17	0.63	-0.02	0.74	0.18
			Polytomous	0.00	0.26	0.01	0.75	-0.02	0.00	0.26	-0.02	0.74	0.02
None	0.2	One	Dichotomous	0.10	0.99	0.01	0.76	0.09	0.11	0.59	0.00	0.74	0.12
			Polytomous	-0.05	0.51	0.01	0.76	-0.06	-0.03	0.35	0.00	0.74	-0.03
		Composite	Dichotomous	-0.05	1.20	0.01	0.76	-0.06	0.20	0.69	0.00	0.74	0.20
			Polytomous	-0.12	0.52	0.01	0.76	-0.13	-0.07	0.34	0.00	0.74	-0.07
		Separate	Dichotomous	0.07	0.71	0.01	0.76	0.06	0.11	0.70	0.00	0.74	0.11
			Polytomous	0.00	0.29	0.01	0.76	-0.01	0.00	0.28	0.00	0.74	0.00
	0.5	One	Dichotomous	0.15	0.63	-0.04	0.77	0.19	0.18	0.59	-0.01	0.76	0.19
			Polytomous	0.00	0.33	-0.04	0.77	0.04	-0.02	0.31	-0.01	0.76	-0.02
		Composite	Dichotomous	0.18	0.61	-0.04	0.77	0.22	0.21	0.58	-0.01	0.76	0.22
			Polytomous	0.00	0.30	-0.04	0.77	0.04	-0.01	0.26	-0.01	0.76	0.00
		Separate	Dichotomous	0.20	0.53	-0.04	0.77	0.24	0.23	0.52	-0.01	0.76	0.23
			Polytomous	0.00	0.23	-0.04	0.77	0.04	0.00	0.21	-0.01	0.76	0.01
	0.7	One	Dichotomous	0.08	0.67	-0.02	0.75	0.09	0.12	0.65	-0.01	0.74	0.13
			Polytomous	-0.02	0.32	-0.02	0.75	0.00	-0.02	0.32	-0.01	0.74	-0.01
		Composite	Dichotomous	0.09	0.65	-0.02	0.75	0.11	0.13	0.64	-0.01	0.74	0.14
			Polytomous	-0.01	0.30	-0.02	0.75	0.00	-0.01	0.29	-0.01	0.74	0.00
		Separate	Dichotomous	0.12	0.62	-0.02	0.75	0.14	0.15	0.61	-0.01	0.74	0.16
			Polytomous	0.00	0.26	-0.02	0.75	0.02	0.00	0.26	-0.01	0.74	0.01
	1.0	One	Dichotomous	0.03	0.71	0.00	0.75	0.03	0.06	0.72	0.01	0.76	0.05

Dependency	Correlation Between Traits	Scaling	Scoring	Percent of CPE									
				30					50				
				Estimated		True		Est.-True	Estimated		True		Est.-True
				mean	sd	mean	sd		mean	sd	mean	sd	
		Composite	Polytomous	-0.03	0.34	0.00	0.75	-0.03	-0.04	0.35	0.01	0.76	-0.06
			Dichotomous	0.05	0.70	0.00	0.75	0.04	0.07	0.73	0.01	0.76	0.06
		Separate	Polytomous	-0.02	0.32	0.00	0.75	-0.03	-0.02	0.32	0.01	0.76	-0.03
			Dichotomous	0.08	0.68	0.00	0.75	0.08	0.09	0.71	0.01	0.76	0.08
			Polytomous	0.00	0.30	0.00	0.75	0.00	0.00	0.30	0.01	0.76	-0.01

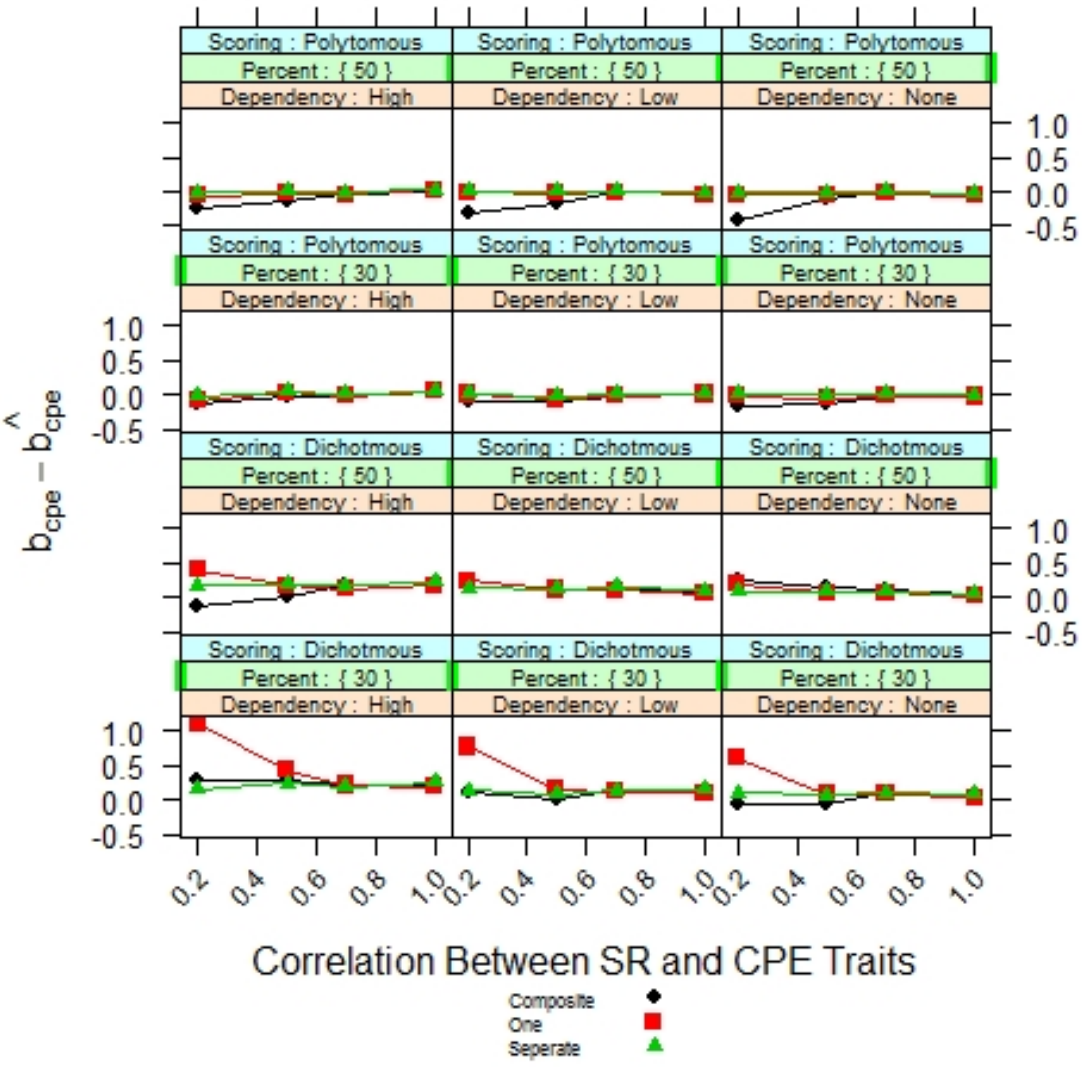

Figure 25: Difference Between Estimated and True CPE "b" Parameters for Sample Size of 3000 and 120 Items

APPENDIX C: RESULTS TABLES AND GRAPHICS

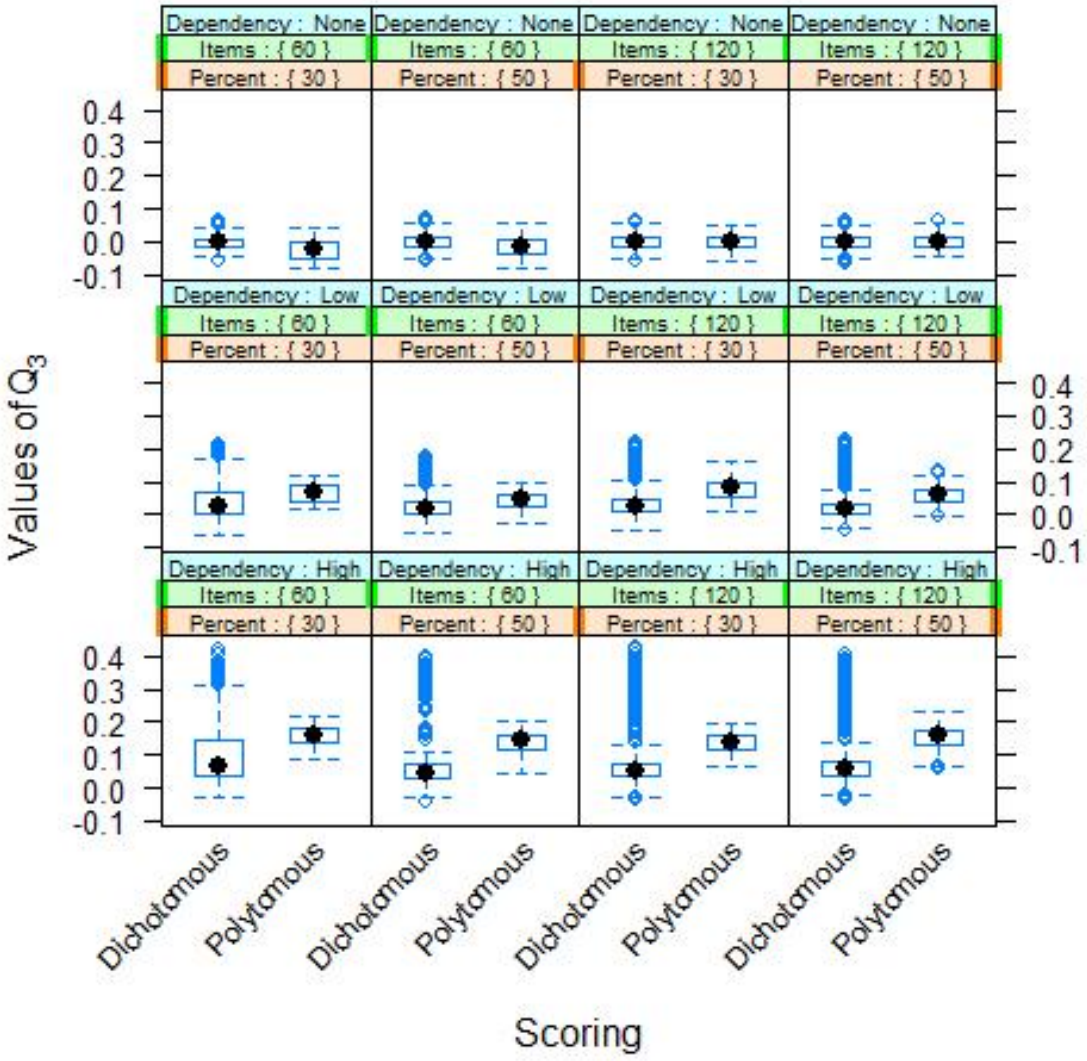

Figure 26: Comparison of Scoring Methods for Sample Size across Different Levels of Dependencies for Sample Size of 1,000

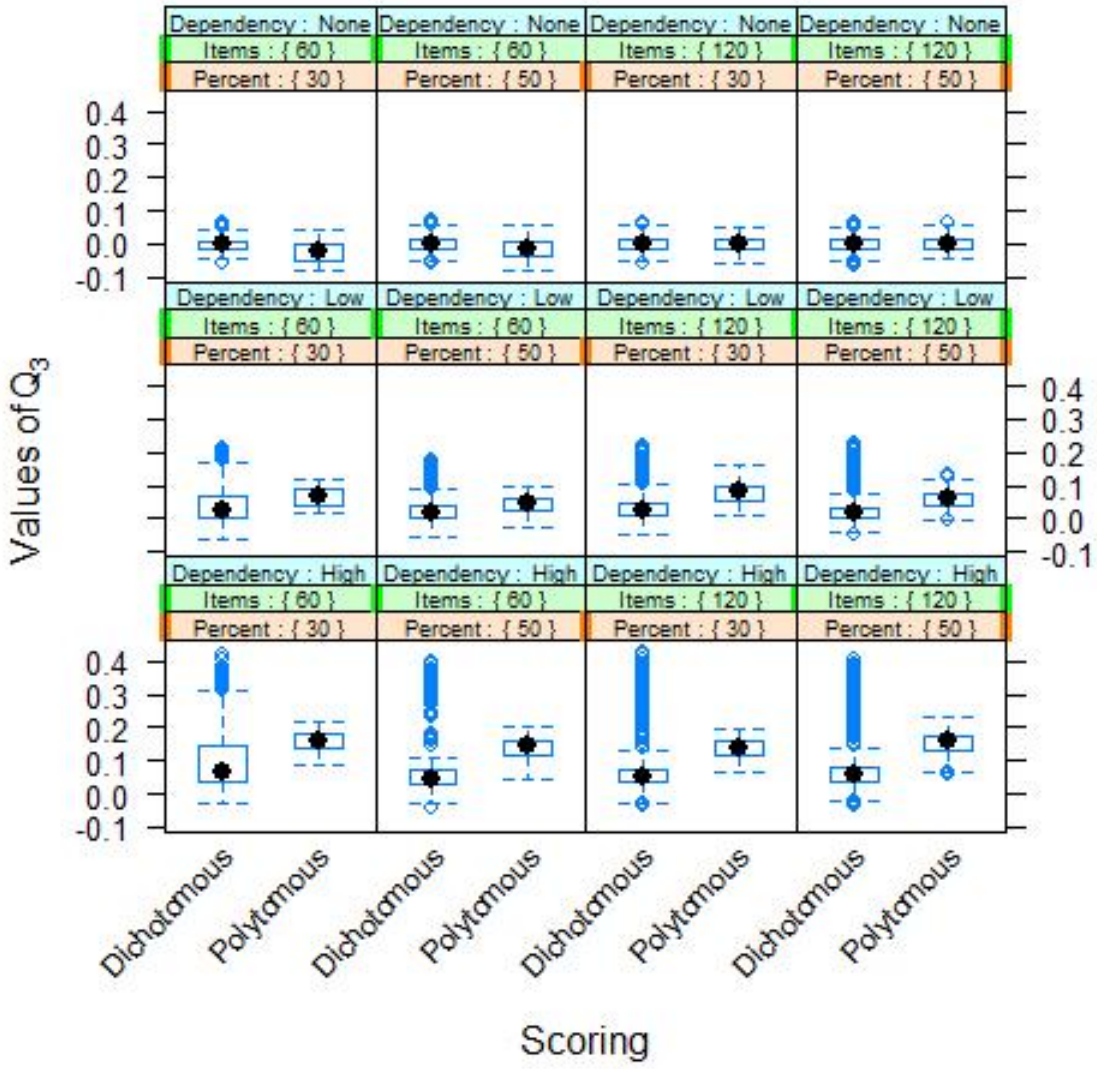

Figure 27: Comparison of Scoring Methods for Sample Size across Different Levels of Dependencies for Sample Size of 3,000

Table 26: Simulation Results for 1,000 Sample Size and 60 Items

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
30				Composite	Between	0.04	0.31	0.19	0.17	0.01	0.04
					Within	0.12	0.57	0.37	0.30	0.01	0.10
			Dichotomous	One	Between	0.04	0.32	0.19	0.18	0.01	0.04
					Within	0.12	0.58	0.37	0.30	0.01	0.10
				Separate	Between	-0.18	0.12	0.18	-0.05	0.01	0.04
					Within	-0.12	0.46	0.47	0.10	0.01	0.13
		0.2		Composite	Between	0.18	0.38	0.10	0.29	0.02	0.03
					Within	0.30	0.52	0.09	0.41	0.02	0.03
			Polytomous	One	Between	0.18	0.38	0.10	0.29	0.02	0.03
					Within	0.30	0.52	0.09	0.41	0.02	0.03
				Separate	Between	-0.19	0.08	0.12	-0.05	0.02	0.04
					Within	-0.02	0.29	0.14	0.15	0.03	0.05
				Composite	Between	-0.01	0.25	0.17	0.12	0.01	0.04
					Within	0.06	0.56	0.39	0.25	0.02	0.11
			Dichotomous	One	Between	0.00	0.26	0.17	0.12	0.01	0.04
					Within	0.06	0.56	0.39	0.25	0.02	0.11
				Separate	Between	-0.19	0.09	0.18	-0.05	0.01	0.04
					Within	-0.13	0.46	0.48	0.10	0.01	0.13
		0.5		Composite	Between	0.12	0.33	0.09	0.24	0.02	0.03
					Within	0.25	0.46	0.09	0.36	0.03	0.03
			Polytomous	One	Between	0.12	0.33	0.09	0.24	0.02	0.03
					Within	0.25	0.46	0.09	0.36	0.02	0.03
				Separate	Between	-0.20	0.05	0.12	-0.05	0.02	0.04
					Within	0.02	0.29	0.13	0.14	0.02	0.05
	High			Composite	Between	-0.06	0.19	0.16	0.07	0.01	0.03
					Within	0.00	0.55	0.41	0.21	0.01	0.11
			Dichotomous	One	Between	-0.06	0.19	0.16	0.07	0.01	0.03
					Within	0.00	0.55	0.41	0.21	0.01	0.11
				Separate	Between	-0.21	0.10	0.18	-0.05	0.02	0.04
					Within	-0.13	0.51	0.47	0.10	0.01	0.13
		0.7		Composite	Between	0.07	0.28	0.10	0.18	0.02	0.03
					Within	0.20	0.40	0.10	0.30	0.02	0.04
			Polytomous	One	Between	0.08	0.28	0.10	0.18	0.02	0.03
					Within	0.20	0.40	0.10	0.30	0.02	0.04
				Separate	Between	-0.17	0.11	0.12	-0.05	0.02	0.04
					Within	-0.02	0.29	0.14	0.14	0.03	0.05
				Composite	Between	-0.14	0.09	0.16	-0.02	0.01	0.03
					Within	-0.06	0.45	0.42	0.13	0.01	0.12
			Dichotomous	One	Between	-0.13	0.10	0.16	-0.02	0.01	0.03
					Within	-0.06	0.45	0.42	0.13	0.01	0.12
				Separate	Between	-0.18	0.10	0.18	-0.05	0.01	0.04
					Within	-0.12	0.45	0.47	0.10	0.01	0.13
		1		Composite	Between	-0.07	0.12	0.10	0.02	0.02	0.03
					Within	0.05	0.26	0.09	0.17	0.02	0.03
			Polytomous	One	Between	-0.07	0.12	0.10	0.02	0.02	0.03
					Within	0.05	0.26	0.09	0.17	0.02	0.03
				Separate	Between	-0.17	0.06	0.10	-0.04	0.01	0.03
					Within	-0.05	0.30	0.14	0.13	0.02	0.05
				Composite	Between	0.06	0.41	0.21	0.24	0.02	0.05
					Within	0.08	0.51	0.27	0.28	0.02	0.06
			Dichotomous	One	Between	0.07	0.42	0.21	0.25	0.02	0.05
					Within	0.08	0.52	0.27	0.29	0.02	0.06
				Separate	Between	-0.17	0.11	0.18	-0.02	0.01	0.04
					Within	-0.10	0.28	0.27	0.04	0.01	0.06
		0.2		Composite	Between	0.32	0.56	0.10	0.45	0.03	0.04
					Within	0.38	0.62	0.09	0.50	0.03	0.04
			Polytomous	One	Between	0.32	0.56	0.10	0.45	0.03	0.04
					Within	0.38	0.62	0.09	0.50	0.03	0.04
				Separate	Between	-0.17	0.09	0.12	-0.03	0.01	0.04
					Within	-0.03	0.19	0.11	0.08	0.02	0.04
				Composite	Between	0.02	0.37	0.22	0.17	0.01	0.04
					Within	0.02	0.46	0.28	0.22	0.02	0.06
		0.5	Dichotomous	One	Between	0.02	0.37	0.22	0.17	0.02	0.04
					Within	0.03	0.46	0.28	0.22	0.02	0.06
				Separate	Between	-0.17	0.10	0.18	-0.02	0.01	0.04
					Within	-0.11	0.26	0.27	0.04	0.01	0.06

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
				Composite	Between	0.24	0.50	0.11	0.36	0.03	0.04
					Within	0.27	0.56	0.10	0.42	0.03	0.04
			Polytomous	One	Between	0.25	0.50	0.11	0.37	0.03	0.04
					Within	0.28	0.56	0.10	0.42	0.03	0.04
				Separate	Between	-0.15	0.08	0.12	-0.03	0.01	0.04
					Within	-0.06	0.21	0.12	0.07	0.02	0.04
				Composite	Between	-0.02	0.23	0.18	0.10	0.01	0.04
					Within	0.00	0.35	0.26	0.15	0.01	0.06
			Dichotomous	One	Between	-0.02	0.23	0.18	0.10	0.01	0.04
					Within	0.00	0.35	0.26	0.16	0.01	0.06
				Separate	Between	-0.15	0.11	0.18	-0.02	0.01	0.04
		0.7			Within	-0.14	0.26	0.26	0.04	0.01	0.06
				Composite	Between	0.13	0.40	0.13	0.26	0.02	0.04
					Within	0.16	0.44	0.11	0.32	0.02	0.04
			Polytomous	One	Between	0.13	0.40	0.13	0.26	0.02	0.04
					Within	0.16	0.44	0.11	0.32	0.02	0.04
				Separate	Between	-0.16	0.08	0.12	-0.03	0.01	0.04
					Within	-0.05	0.22	0.13	0.08	0.02	0.05
				Composite	Between	-0.12	0.11	0.16	-0.01	0.01	0.03
					Within	-0.12	0.30	0.27	0.05	0.01	0.06
			Dichotomous	One	Between	-0.12	0.11	0.16	-0.01	0.01	0.03
					Within	-0.12	0.30	0.27	0.05	0.01	0.06
				Separate	Between	-0.15	0.10	0.18	-0.02	0.01	0.04
		1			Within	-0.11	0.29	0.27	0.04	0.01	0.06
				Composite	Between	-0.12	0.07	0.09	-0.02	0.01	0.03
					Within	-0.06	0.25	0.10	0.07	0.03	0.04
			Polytomous	One	Between	-0.12	0.07	0.09	-0.02	0.01	0.03
					Within	-0.06	0.25	0.10	0.07	0.03	0.04
				Separate	Between	-0.14	0.10	0.11	-0.03	0.02	0.04
					Within	-0.03	0.20	0.11	0.08	0.02	0.04
	None			Composite	Between	0.06	0.52	0.27	0.27	0.02	0.06
					Within	0.06	0.51	0.26	0.27	0.02	0.06
			Dichotomous	One	Between	0.06	0.53	0.27	0.28	0.02	0.06
					Within	0.07	0.52	0.27	0.28	0.02	0.06
				Separate	Between	-0.13	0.13	0.18	0.01	0.00	0.04
		0.2			Within	-0.15	0.17	0.18	0.01	0.00	0.04
				Composite	Between	0.40	0.66	0.10	0.56	0.03	0.03
					Within	0.41	0.64	0.09	0.56	0.02	0.03
			Polytomous	One	Between	0.41	0.67	0.10	0.56	0.03	0.03
					Within	0.41	0.65	0.09	0.56	0.03	0.03
				Separate	Between	-0.10	0.14	0.12	0.01	0.01	0.04
					Within	-0.13	0.11	0.11	0.00	0.01	0.04
				Composite	Between	0.01	0.38	0.23	0.20	0.01	0.05
					Within	0.01	0.39	0.23	0.20	0.01	0.05
			Dichotomous	One	Between	0.01	0.38	0.23	0.20	0.01	0.05
					Within	0.01	0.38	0.23	0.20	0.01	0.05
				Separate	Between	-0.14	0.13	0.18	0.01	0.00	0.04
		0.5			Within	-0.13	0.13	0.17	0.01	0.00	0.04
				Composite	Between	0.29	0.58	0.12	0.46	0.02	0.04
					Within	0.31	0.56	0.11	0.46	0.02	0.04
			Polytomous	One	Between	0.30	0.58	0.12	0.46	0.02	0.04
					Within	0.31	0.57	0.11	0.46	0.02	0.04
				Separate	Between	-0.12	0.11	0.12	0.01	0.01	0.04
					Within	-0.12	0.11	0.10	0.01	0.02	0.04
				Composite	Between	0.01	0.26	0.19	0.12	0.01	0.04
					Within	-0.04	0.27	0.19	0.12	0.01	0.04
			Dichotomous	One	Between	0.01	0.26	0.19	0.12	0.01	0.04
					Within	-0.04	0.27	0.19	0.12	0.01	0.04
				Separate	Between	-0.13	0.13	0.18	0.01	0.00	0.04
		0.7			Within	-0.13	0.12	0.17	0.01	0.00	0.04
				Composite	Between	0.18	0.46	0.13	0.33	0.02	0.04
					Within	0.21	0.42	0.11	0.32	0.02	0.04
			Polytomous	One	Between	0.19	0.46	0.13	0.33	0.02	0.04
					Within	0.21	0.42	0.11	0.33	0.02	0.04
				Separate	Between	-0.09	0.11	0.12	0.01	0.01	0.04
					Within	-0.10	0.14	0.10	0.00	0.01	0.04
		1	Dichotomous	Composite	Between	-0.13	0.10	0.16	0.00	0.00	0.03
					Within	-0.14	0.12	0.16	0.00	0.00	0.03
				One	Between	-0.13	0.10	0.16	0.00	0.00	0.03

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
					Within	-0.14	0.12	0.16	0.00	0.00	0.03
				Separate	Between	-0.13	0.12	0.18	0.01	0.00	0.04
					Within	-0.15	0.12	0.17	0.01	0.00	0.04
			Polytomous	Composite	Between	-0.13	0.06	0.10	-0.03	0.01	0.03
					Within	-0.13	0.06	0.08	-0.03	0.01	0.03
				One	Between	-0.13	0.06	0.10	-0.03	0.01	0.03
					Within	-0.13	0.06	0.08	-0.03	0.01	0.03
				Separate	Between	-0.12	0.12	0.11	0.00	0.01	0.04
					Within	-0.13	0.14	0.11	0.01	0.01	0.04
50			Dichotomous	Composite	Between	-0.10	0.28	0.21	0.09	0.04	0.04
					Within	-0.02	0.54	0.43	0.20	0.03	0.09
				One	Between	-0.03	0.34	0.21	0.14	0.02	0.04
					Within	0.05	0.57	0.42	0.24	0.02	0.09
				Separate	Between	-0.18	0.09	0.19	-0.05	0.01	0.03
		0.2			Within	-0.09	0.48	0.46	0.07	0.01	0.10
			Polytomous	Composite	Between	0.12	0.40	0.13	0.28	0.02	0.03
					Within	0.27	0.51	0.13	0.39	0.02	0.03
				One	Between	0.15	0.40	0.13	0.29	0.01	0.03
					Within	0.28	0.51	0.13	0.40	0.01	0.03
				Separate	Between	-0.17	0.07	0.13	-0.04	0.01	0.03
					Within	0.01	0.24	0.14	0.13	0.02	0.04
			Dichotomous	Composite	Between	-0.11	0.18	0.19	0.05	0.02	0.03
					Within	-0.04	0.52	0.43	0.16	0.02	0.09
				One	Between	-0.08	0.21	0.19	0.06	0.01	0.03
					Within	0.00	0.52	0.43	0.17	0.01	0.09
				Separate	Between	-0.17	0.09	0.19	-0.05	0.01	0.04
		0.5			Within	-0.14	0.45	0.47	0.07	0.01	0.10
			Polytomous	Composite	Between	0.06	0.31	0.14	0.20	0.02	0.04
					Within	0.16	0.45	0.14	0.33	0.02	0.04
				One	Between	0.08	0.32	0.14	0.22	0.02	0.04
					Within	0.19	0.45	0.14	0.34	0.02	0.04
				Separate	Between	-0.17	0.07	0.14	-0.04	0.01	0.03
	High				Within	0.00	0.29	0.15	0.13	0.02	0.04
			Dichotomous	Composite	Between	-0.11	0.16	0.18	0.01	0.01	0.03
					Within	-0.02	0.49	0.44	0.13	0.01	0.09
				One	Between	-0.10	0.16	0.18	0.02	0.01	0.03
					Within	-0.02	0.49	0.44	0.14	0.01	0.09
				Separate	Between	-0.18	0.07	0.19	-0.05	0.01	0.03
		0.7			Within	-0.10	0.47	0.48	0.07	0.01	0.10
			Polytomous	Composite	Between	0.01	0.25	0.13	0.14	0.02	0.03
					Within	0.14	0.41	0.13	0.27	0.02	0.04
				One	Between	0.01	0.25	0.13	0.14	0.02	0.03
					Within	0.14	0.41	0.13	0.27	0.02	0.04
				Separate	Between	-0.16	0.07	0.13	-0.04	0.01	0.03
					Within	0.01	0.28	0.15	0.13	0.01	0.04
			Dichotomous	Composite	Between	-0.15	0.08	0.18	-0.04	0.01	0.03
					Within	-0.10	0.46	0.45	0.09	0.01	0.10
				One	Between	-0.15	0.08	0.18	-0.04	0.00	0.03
					Within	-0.10	0.46	0.45	0.09	0.01	0.10
				Separate	Between	-0.17	0.07	0.19	-0.05	0.01	0.03
		1			Within	-0.11	0.44	0.47	0.07	0.01	0.10
			Polytomous	Composite	Between	-0.12	0.11	0.13	0.00	0.01	0.03
					Within	0.03	0.27	0.15	0.15	0.02	0.04
				One	Between	-0.13	0.10	0.13	0.00	0.01	0.03
					Within	0.03	0.26	0.15	0.15	0.02	0.04
				Separate	Between	-0.15	0.08	0.13	-0.04	0.01	0.03
					Within	0.00	0.25	0.16	0.13	0.01	0.04
	Low		Dichotomous	Composite	Between	-0.12	0.32	0.24	0.06	0.04	0.04
					Within	-0.10	0.36	0.30	0.10	0.04	0.05
				One	Between	-0.05	0.36	0.23	0.16	0.03	0.04
					Within	-0.02	0.44	0.29	0.19	0.03	0.05
				Separate	Between	-0.14	0.12	0.19	-0.02	0.01	0.03
		0.2			Within	-0.12	0.28	0.28	0.03	0.01	0.05
			Polytomous	Composite	Between	0.23	0.56	0.15	0.40	0.03	0.04
					Within	0.29	0.60	0.14	0.45	0.03	0.04
				One	Between	0.29	0.59	0.14	0.44	0.02	0.04
					Within	0.34	0.62	0.14	0.48	0.02	0.04
				Separate	Between	-0.15	0.08	0.13	-0.02	0.01	0.03
					Within	-0.06	0.19	0.14	0.07	0.02	0.04

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
				Composite	Between	-0.11	0.24	0.21	0.05	0.02	0.04
					Within	-0.09	0.34	0.30	0.09	0.02	0.05
			Dichotomous	One	Between	-0.07	0.24	0.20	0.07	0.01	0.04
					Within	-0.06	0.35	0.29	0.12	0.01	0.05
				Separate	Between	-0.16	0.12	0.19	-0.02	0.01	0.04
		0.5			Within	-0.12	0.28	0.29	0.03	0.01	0.05
				Composite	Between	0.10	0.44	0.17	0.27	0.03	0.04
					Within	0.16	0.49	0.16	0.33	0.03	0.04
			Polytomous	One	Between	0.13	0.44	0.16	0.28	0.03	0.04
					Within	0.18	0.49	0.15	0.34	0.03	0.04
				Separate	Between	-0.14	0.08	0.13	-0.03	0.01	0.03
					Within	-0.05	0.20	0.13	0.07	0.02	0.04
				Composite	Between	-0.09	0.17	0.19	0.03	0.01	0.03
					Within	-0.06	0.31	0.27	0.08	0.01	0.05
			Dichotomous	One	Between	-0.09	0.17	0.19	0.04	0.01	0.03
					Within	-0.05	0.32	0.27	0.08	0.01	0.05
				Separate	Between	-0.16	0.10	0.19	-0.02	0.01	0.04
		0.7			Within	-0.11	0.29	0.29	0.03	0.01	0.05
				Composite	Between	-0.01	0.30	0.16	0.15	0.02	0.04
					Within	0.06	0.37	0.15	0.22	0.03	0.04
			Polytomous	One	Between	0.00	0.31	0.16	0.16	0.02	0.04
					Within	0.07	0.37	0.15	0.23	0.03	0.04
				Separate	Between	-0.13	0.10	0.14	-0.03	0.01	0.04
					Within	-0.04	0.19	0.13	0.07	0.01	0.04
				Composite	Between	-0.14	0.12	0.18	-0.02	0.00	0.03
					Within	-0.11	0.25	0.28	0.03	0.01	0.05
			Dichotomous	One	Between	-0.14	0.12	0.18	-0.02	0.00	0.03
					Within	-0.12	0.25	0.28	0.03	0.01	0.05
				Separate	Between	-0.15	0.13	0.19	-0.02	0.01	0.03
		1			Within	-0.10	0.25	0.28	0.03	0.01	0.05
				Composite	Between	-0.15	0.06	0.13	-0.03	0.01	0.03
					Within	-0.06	0.18	0.13	0.06	0.02	0.03
			Polytomous	One	Between	-0.15	0.06	0.13	-0.03	0.01	0.03
					Within	-0.06	0.18	0.13	0.06	0.02	0.03
				Separate	Between	-0.16	0.09	0.13	-0.03	0.01	0.03
					Within	-0.04	0.22	0.13	0.07	0.02	0.03
	None			Composite	Between	-0.12	0.33	0.25	0.04	0.03	0.04
					Within	-0.11	0.28	0.24	0.04	0.03	0.04
			Dichotomous	One	Between	-0.05	0.38	0.24	0.13	0.04	0.04
					Within	-0.03	0.39	0.24	0.13	0.04	0.04
				Separate	Between	-0.13	0.13	0.19	0.01	0.00	0.03
		0.2			Within	-0.15	0.13	0.19	0.01	0.00	0.03
				Composite	Between	-0.06	0.62	0.18	0.45	0.08	0.05
					Within	-0.06	0.60	0.17	0.44	0.08	0.05
			Polytomous	One	Between	0.33	0.64	0.16	0.52	0.03	0.04
					Within	0.36	0.64	0.16	0.52	0.03	0.04
				Separate	Between	-0.11	0.11	0.15	0.01	0.01	0.04
					Within	-0.10	0.11	0.12	0.01	0.01	0.03
				Composite	Between	-0.10	0.26	0.22	0.06	0.02	0.04
					Within	-0.09	0.25	0.22	0.06	0.02	0.04
			Dichotomous	One	Between	-0.07	0.25	0.21	0.09	0.01	0.04
					Within	-0.04	0.25	0.21	0.09	0.01	0.04
				Separate	Between	-0.15	0.13	0.19	0.01	0.00	0.03
		0.5			Within	-0.16	0.13	0.20	0.01	0.00	0.03
				Composite	Between	0.01	0.52	0.18	0.27	0.06	0.05
					Within	0.05	0.48	0.19	0.27	0.05	0.05
			Polytomous	One	Between	0.11	0.49	0.18	0.31	0.03	0.05
					Within	0.13	0.46	0.18	0.31	0.03	0.05
				Separate	Between	-0.09	0.14	0.14	0.01	0.01	0.03
					Within	-0.10	0.12	0.13	0.01	0.01	0.04
				Composite	Between	-0.08	0.18	0.19	0.04	0.01	0.03
		0.7			Within	-0.09	0.18	0.19	0.04	0.01	0.03
			Dichotomous	One	Between	-0.08	0.19	0.20	0.05	0.01	0.03
					Within	-0.08	0.20	0.19	0.05	0.01	0.03
				Separate	Between	-0.12	0.13	0.19	0.01	0.00	0.03
					Within	-0.15	0.13	0.19	0.00	0.00	0.03
				Composite	Between	0.01	0.29	0.15	0.14	0.02	0.04
			Polytomous		Within	0.00	0.33	0.15	0.14	0.03	0.04
				One	Between	0.02	0.30	0.15	0.16	0.02	0.04

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
					Within	0.03	0.35	0.15	0.17	0.02	0.04
				Separate	Between	-0.09	0.12	0.13	0.01	0.01	0.03
					Within	-0.10	0.12	0.13	0.01	0.01	0.03
			Dichotomous	Composite	Between	-0.14	0.12	0.19	0.00	0.00	0.03
					Within	-0.12	0.13	0.18	0.00	0.00	0.03
				One	Between	-0.14	0.12	0.19	0.00	0.00	0.03
					Within	-0.12	0.12	0.18	0.00	0.00	0.03
				Separate	Between	-0.14	0.12	0.19	0.00	0.00	0.03
		1			Within	-0.14	0.13	0.19	0.01	0.00	0.03
			Polytomous	Composite	Between	-0.12	0.07	0.12	-0.02	0.01	0.03
					Within	-0.12	0.07	0.13	-0.02	0.01	0.03
				One	Between	-0.12	0.07	0.12	-0.02	0.01	0.03
					Within	-0.12	0.07	0.13	-0.02	0.01	0.03
				Separate	Between	-0.10	0.11	0.13	0.01	0.00	0.03
					Within	-0.10	0.10	0.13	0.01	0.01	0.03

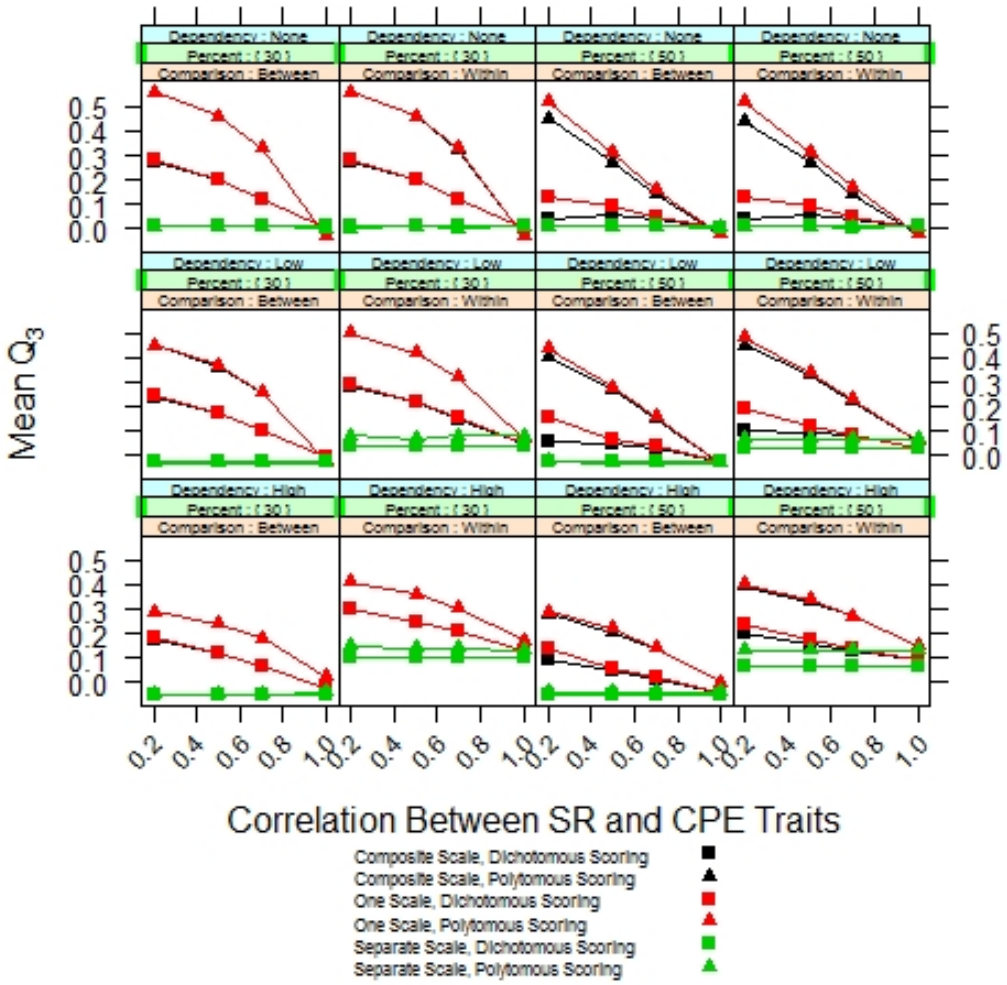

Figure 28: Average Q3 for Sample Size of 1000 and 60 Items

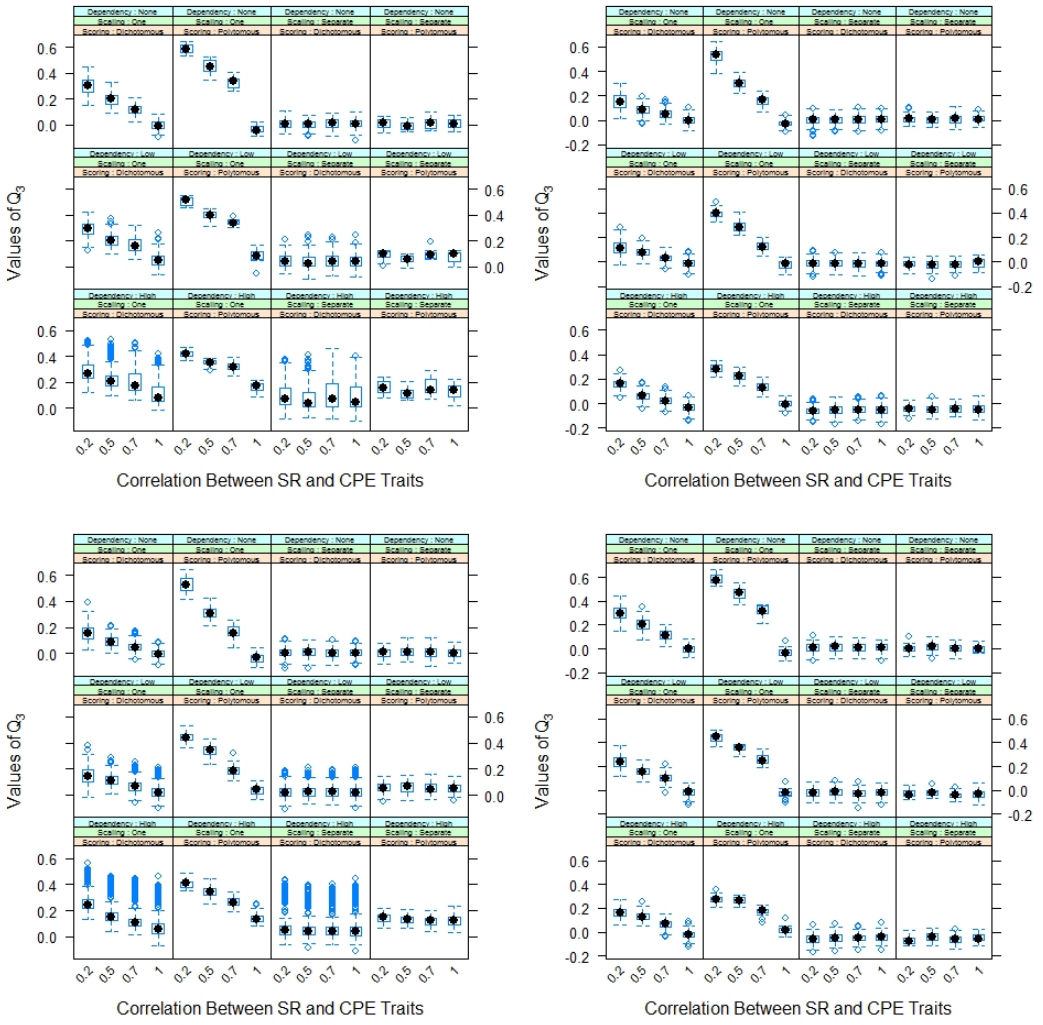

Figure 29: A-D 1000 and 60 Items for Within and Between Comparisons for 30% and 50% CPEs

Note: Clockwise from Top Left: 1) Within, 30%, 2) Between, 50%, 3) Within, 50%, 4) Between, 50%

Table 27: Simulation Results for 1,000 Sample Size and 120 Items

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
30				Composite	Between	0.04	0.31	0.19	0.17	0.01	0.04
					Within	0.12	0.57	0.37	0.30	0.01	0.10
			Dichotomous	One	Between	0.04	0.32	0.19	0.18	0.01	0.04
					Within	0.12	0.58	0.37	0.30	0.01	0.10
				Separate	Between	-0.18	0.12	0.18	-0.05	0.01	0.04
					Within	-0.12	0.46	0.47	0.10	0.01	0.13
		0.2		Composite	Between	0.18	0.38	0.10	0.29	0.02	0.03
					Within	0.30	0.52	0.09	0.41	0.02	0.03
			Polytomous	One	Between	0.18	0.38	0.10	0.29	0.02	0.03
					Within	0.30	0.52	0.09	0.41	0.02	0.03
				Separate	Between	-0.19	0.08	0.12	-0.05	0.02	0.04
					Within	-0.02	0.29	0.14	0.15	0.03	0.05
				Composite	Between	-0.01	0.25	0.17	0.12	0.01	0.04
					Within	0.06	0.56	0.39	0.25	0.02	0.11
			Dichotomous	One	Between	0.00	0.26	0.17	0.12	0.01	0.04
					Within	0.06	0.56	0.39	0.25	0.02	0.11
				Separate	Between	-0.19	0.09	0.18	-0.05	0.01	0.04
					Within	-0.13	0.46	0.48	0.10	0.01	0.13
		0.5		Composite	Between	0.12	0.33	0.09	0.24	0.02	0.03
					Within	0.25	0.46	0.09	0.36	0.03	0.03
			Polytomous	One	Between	0.12	0.33	0.09	0.24	0.02	0.03
					Within	0.25	0.46	0.09	0.36	0.02	0.03
				Separate	Between	-0.20	0.05	0.12	-0.05	0.02	0.04
					Within	0.02	0.29	0.13	0.14	0.02	0.05
	High			Composite	Between	-0.06	0.19	0.16	0.07	0.01	0.03
					Within	0.00	0.55	0.41	0.21	0.01	0.11
			Dichotomous	One	Between	-0.06	0.19	0.16	0.07	0.01	0.03
					Within	0.00	0.55	0.41	0.21	0.01	0.11
				Separate	Between	-0.21	0.10	0.18	-0.05	0.02	0.04
					Within	-0.13	0.51	0.47	0.10	0.01	0.13
		0.7		Composite	Between	0.07	0.28	0.10	0.18	0.02	0.03
					Within	0.20	0.40	0.10	0.30	0.02	0.04
			Polytomous	One	Between	0.08	0.28	0.10	0.18	0.02	0.03
					Within	0.20	0.40	0.10	0.30	0.02	0.04
				Separate	Between	-0.17	0.11	0.12	-0.05	0.02	0.04
					Within	-0.02	0.29	0.14	0.14	0.03	0.05
				Composite	Between	-0.14	0.09	0.16	-0.02	0.01	0.03
					Within	-0.06	0.45	0.42	0.13	0.01	0.12
			Dichotomous	One	Between	-0.13	0.10	0.16	-0.02	0.01	0.03
					Within	-0.06	0.45	0.42	0.13	0.01	0.12
				Separate	Between	-0.18	0.10	0.18	-0.05	0.01	0.04
					Within	-0.12	0.45	0.47	0.10	0.01	0.13
		1		Composite	Between	-0.07	0.12	0.10	0.02	0.02	0.03
					Within	0.05	0.26	0.09	0.17	0.02	0.03
			Polytomous	One	Between	-0.07	0.12	0.10	0.02	0.02	0.03
					Within	0.05	0.26	0.09	0.17	0.02	0.03
				Separate	Between	-0.17	0.06	0.10	-0.04	0.01	0.03
					Within	-0.05	0.30	0.14	0.13	0.02	0.05
				Composite	Between	0.06	0.41	0.21	0.24	0.02	0.05
					Within	0.08	0.51	0.27	0.28	0.02	0.06
			Dichotomous	One	Between	0.07	0.42	0.21	0.25	0.02	0.05
					Within	0.08	0.52	0.27	0.29	0.02	0.06
				Separate	Between	-0.17	0.11	0.18	-0.02	0.01	0.04
					Within	-0.10	0.28	0.27	0.04	0.01	0.06
		0.2		Composite	Between	0.32	0.56	0.10	0.45	0.03	0.04
					Within	0.38	0.62	0.09	0.50	0.03	0.04
			Polytomous	One	Between	0.32	0.56	0.10	0.45	0.03	0.04
					Within	0.38	0.62	0.09	0.50	0.03	0.04
				Separate	Between	-0.17	0.09	0.12	-0.03	0.01	0.04
					Within	-0.03	0.19	0.11	0.08	0.02	0.04
				Composite	Between	0.02	0.37	0.22	0.17	0.01	0.04
					Within	0.02	0.46	0.28	0.22	0.02	0.06
		0.5	Dichotomous	One	Between	0.02	0.37	0.22	0.17	0.02	0.04
					Within	0.03	0.46	0.28	0.22	0.02	0.06
				Separate	Between	-0.17	0.10	0.18	-0.02	0.01	0.04
					Within	-0.11	0.26	0.27	0.04	0.01	0.06

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
				Composite	Between	0.24	0.50	0.11	0.36	0.03	0.04
					Within	0.27	0.56	0.10	0.42	0.03	0.04
			Polytomous	One	Between	0.25	0.50	0.11	0.37	0.03	0.04
					Within	0.28	0.56	0.10	0.42	0.03	0.04
				Separate	Between	-0.15	0.08	0.12	-0.03	0.01	0.04
					Within	-0.06	0.21	0.12	0.07	0.02	0.04
				Composite	Between	-0.02	0.23	0.18	0.10	0.01	0.04
					Within	0.00	0.35	0.26	0.15	0.01	0.06
			Dichotomous	One	Between	-0.02	0.23	0.18	0.10	0.01	0.04
					Within	0.00	0.35	0.26	0.16	0.01	0.06
				Separate	Between	-0.15	0.11	0.18	-0.02	0.01	0.04
		0.7			Within	-0.14	0.26	0.26	0.04	0.01	0.06
				Composite	Between	0.13	0.40	0.13	0.26	0.02	0.04
					Within	0.16	0.44	0.11	0.32	0.02	0.04
			Polytomous	One	Between	0.13	0.40	0.13	0.26	0.02	0.04
					Within	0.16	0.44	0.11	0.32	0.02	0.04
				Separate	Between	-0.16	0.08	0.12	-0.03	0.01	0.04
					Within	-0.05	0.22	0.13	0.08	0.02	0.05
				Composite	Between	-0.12	0.11	0.16	-0.01	0.01	0.03
					Within	-0.12	0.30	0.27	0.05	0.01	0.06
			Dichotomous	One	Between	-0.12	0.11	0.16	-0.01	0.01	0.03
					Within	-0.12	0.30	0.27	0.05	0.01	0.06
				Separate	Between	-0.15	0.10	0.18	-0.02	0.01	0.04
		1			Within	-0.11	0.29	0.27	0.04	0.01	0.06
				Composite	Between	-0.12	0.07	0.09	-0.02	0.01	0.03
					Within	-0.06	0.25	0.10	0.07	0.03	0.04
			Polytomous	One	Between	-0.12	0.07	0.09	-0.02	0.01	0.03
					Within	-0.06	0.25	0.10	0.07	0.03	0.04
				Separate	Between	-0.14	0.10	0.11	-0.03	0.02	0.04
					Within	-0.03	0.20	0.11	0.08	0.02	0.04
	None			Composite	Between	0.06	0.52	0.27	0.27	0.02	0.06
					Within	0.06	0.51	0.26	0.27	0.02	0.06
			Dichotomous	One	Between	0.06	0.53	0.27	0.28	0.02	0.06
					Within	0.07	0.52	0.27	0.28	0.02	0.06
				Separate	Between	-0.13	0.13	0.18	0.01	0.00	0.04
		0.2			Within	-0.15	0.17	0.18	0.01	0.00	0.04
				Composite	Between	0.40	0.66	0.10	0.56	0.03	0.03
					Within	0.41	0.64	0.09	0.56	0.02	0.03
			Polytomous	One	Between	0.41	0.67	0.10	0.56	0.03	0.03
					Within	0.41	0.65	0.09	0.56	0.03	0.03
				Separate	Between	-0.10	0.14	0.12	0.01	0.01	0.04
					Within	-0.13	0.11	0.11	0.00	0.01	0.04
				Composite	Between	0.01	0.38	0.23	0.20	0.01	0.05
					Within	0.01	0.39	0.23	0.20	0.01	0.05
			Dichotomous	One	Between	0.01	0.38	0.23	0.20	0.01	0.05
					Within	0.01	0.38	0.23	0.20	0.01	0.05
				Separate	Between	-0.14	0.13	0.18	0.01	0.00	0.04
		0.5			Within	-0.13	0.13	0.17	0.01	0.00	0.04
				Composite	Between	0.29	0.58	0.12	0.46	0.02	0.04
					Within	0.31	0.56	0.11	0.46	0.02	0.04
			Polytomous	One	Between	0.30	0.58	0.12	0.46	0.02	0.04
					Within	0.31	0.57	0.11	0.46	0.02	0.04
				Separate	Between	-0.12	0.11	0.12	0.01	0.01	0.04
					Within	-0.12	0.11	0.10	0.01	0.02	0.04
				Composite	Between	0.01	0.26	0.19	0.12	0.01	0.04
					Within	-0.04	0.27	0.19	0.12	0.01	0.04
			Dichotomous	One	Between	0.01	0.26	0.19	0.12	0.01	0.04
					Within	-0.04	0.27	0.19	0.12	0.01	0.04
				Separate	Between	-0.13	0.13	0.18	0.01	0.00	0.04
		0.7			Within	-0.13	0.12	0.17	0.01	0.00	0.04
				Composite	Between	0.18	0.46	0.13	0.33	0.02	0.04
					Within	0.21	0.42	0.11	0.32	0.02	0.04
			Polytomous	One	Between	0.19	0.46	0.13	0.33	0.02	0.04
					Within	0.21	0.42	0.11	0.33	0.02	0.04
				Separate	Between	-0.09	0.11	0.12	0.01	0.01	0.04
					Within	-0.10	0.14	0.10	0.00	0.01	0.04
		1	Dichotomous	Composite	Between	-0.13	0.10	0.16	0.00	0.00	0.03
					Within	-0.14	0.12	0.16	0.00	0.00	0.03
				One	Between	-0.13	0.10	0.16	0.00	0.00	0.03

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
					Within	-0.14	0.12	0.16	0.00	0.00	0.03
				Separate	Between	-0.13	0.12	0.18	0.01	0.00	0.04
					Within	-0.15	0.12	0.17	0.01	0.00	0.04
			Polytomous	Composite	Between	-0.13	0.06	0.10	-0.03	0.01	0.03
					Within	-0.13	0.06	0.08	-0.03	0.01	0.03
				One	Between	-0.13	0.06	0.10	-0.03	0.01	0.03
					Within	-0.13	0.06	0.08	-0.03	0.01	0.03
				Separate	Between	-0.12	0.12	0.11	0.00	0.01	0.04
					Within	-0.13	0.14	0.11	0.01	0.01	0.04
50				Composite	Between	-0.10	0.28	0.21	0.09	0.04	0.04
					Within	-0.02	0.54	0.43	0.20	0.03	0.09
			Dichotomous	One	Between	-0.03	0.34	0.21	0.14	0.02	0.04
					Within	0.05	0.57	0.42	0.24	0.02	0.09
				Separate	Between	-0.18	0.09	0.19	-0.05	0.01	0.03
		0.2			Within	-0.09	0.48	0.46	0.07	0.01	0.10
				Composite	Between	0.12	0.40	0.13	0.28	0.02	0.03
					Within	0.27	0.51	0.13	0.39	0.02	0.03
			Polytomous	One	Between	0.15	0.40	0.13	0.29	0.01	0.03
					Within	0.28	0.51	0.13	0.40	0.01	0.03
				Separate	Between	-0.17	0.07	0.13	-0.04	0.01	0.03
					Within	0.01	0.24	0.14	0.13	0.02	0.04
				Composite	Between	-0.11	0.18	0.19	0.05	0.02	0.03
					Within	-0.04	0.52	0.43	0.16	0.02	0.09
			Dichotomous	One	Between	-0.08	0.21	0.19	0.06	0.01	0.03
					Within	0.00	0.52	0.43	0.17	0.01	0.09
				Separate	Between	-0.17	0.09	0.19	-0.05	0.01	0.04
		0.5			Within	-0.14	0.45	0.47	0.07	0.01	0.10
				Composite	Between	0.06	0.31	0.14	0.20	0.02	0.04
					Within	0.16	0.45	0.14	0.33	0.02	0.04
			Polytomous	One	Between	0.08	0.32	0.14	0.22	0.02	0.04
					Within	0.19	0.45	0.14	0.34	0.02	0.04
				Separate	Between	-0.17	0.07	0.14	-0.04	0.01	0.03
	High				Within	0.00	0.29	0.15	0.13	0.02	0.04
				Composite	Between	-0.11	0.16	0.18	0.01	0.01	0.03
					Within	-0.02	0.49	0.44	0.13	0.01	0.09
			Dichotomous	One	Between	-0.10	0.16	0.18	0.02	0.01	0.03
					Within	-0.02	0.49	0.44	0.14	0.01	0.09
				Separate	Between	-0.18	0.07	0.19	-0.05	0.01	0.03
		0.7			Within	-0.10	0.47	0.48	0.07	0.01	0.10
				Composite	Between	0.01	0.25	0.13	0.14	0.02	0.03
					Within	0.14	0.41	0.13	0.27	0.02	0.04
			Polytomous	One	Between	0.01	0.25	0.13	0.14	0.02	0.03
					Within	0.14	0.41	0.13	0.27	0.02	0.04
				Separate	Between	-0.16	0.07	0.13	-0.04	0.01	0.03
					Within	0.01	0.28	0.15	0.13	0.01	0.04
				Composite	Between	-0.15	0.08	0.18	-0.04	0.01	0.03
					Within	-0.10	0.46	0.45	0.09	0.01	0.10
			Dichotomous	One	Between	-0.15	0.08	0.18	-0.04	0.00	0.03
					Within	-0.10	0.46	0.45	0.09	0.01	0.10
				Separate	Between	-0.17	0.07	0.19	-0.05	0.01	0.03
		1			Within	-0.11	0.44	0.47	0.07	0.01	0.10
				Composite	Between	-0.12	0.11	0.13	0.00	0.01	0.03
					Within	0.03	0.27	0.15	0.15	0.02	0.04
			Polytomous	One	Between	-0.13	0.10	0.13	0.00	0.01	0.03
					Within	0.03	0.26	0.15	0.15	0.02	0.04
				Separate	Between	-0.15	0.08	0.13	-0.04	0.01	0.03
					Within	0.00	0.25	0.16	0.13	0.01	0.04
	Low			Composite	Between	-0.12	0.32	0.24	0.06	0.04	0.04
					Within	-0.10	0.36	0.30	0.10	0.04	0.05
			Dichotomous	One	Between	-0.05	0.36	0.23	0.16	0.03	0.04
					Within	-0.02	0.44	0.29	0.19	0.03	0.05
				Separate	Between	-0.14	0.12	0.19	-0.02	0.01	0.03
		0.2			Within	-0.12	0.28	0.28	0.03	0.01	0.05
				Composite	Between	0.23	0.56	0.15	0.40	0.03	0.04
					Within	0.29	0.60	0.14	0.45	0.03	0.04
			Polytomous	One	Between	0.29	0.59	0.14	0.44	0.02	0.04
					Within	0.34	0.62	0.14	0.48	0.02	0.04
				Separate	Between	-0.15	0.08	0.13	-0.02	0.01	0.03
					Within	-0.06	0.19	0.14	0.07	0.02	0.04

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
				Composite	Between	-0.11	0.24	0.21	0.05	0.02	0.04
					Within	-0.09	0.34	0.30	0.09	0.02	0.05
			Dichotomous	One	Between	-0.07	0.24	0.20	0.07	0.01	0.04
					Within	-0.06	0.35	0.29	0.12	0.01	0.05
				Separate	Between	-0.16	0.12	0.19	-0.02	0.01	0.04
		0.5			Within	-0.12	0.28	0.29	0.03	0.01	0.05
				Composite	Between	0.10	0.44	0.17	0.27	0.03	0.04
					Within	0.16	0.49	0.16	0.33	0.03	0.04
			Polytomous	One	Between	0.13	0.44	0.16	0.28	0.03	0.04
					Within	0.18	0.49	0.15	0.34	0.03	0.04
				Separate	Between	-0.14	0.08	0.13	-0.03	0.01	0.03
					Within	-0.05	0.20	0.13	0.07	0.02	0.04
				Composite	Between	-0.09	0.17	0.19	0.03	0.01	0.03
					Within	-0.06	0.31	0.27	0.08	0.01	0.05
			Dichotomous	One	Between	-0.09	0.17	0.19	0.04	0.01	0.03
					Within	-0.05	0.32	0.27	0.08	0.01	0.05
				Separate	Between	-0.16	0.10	0.19	-0.02	0.01	0.04
		0.7			Within	-0.11	0.29	0.29	0.03	0.01	0.05
				Composite	Between	-0.01	0.30	0.16	0.15	0.02	0.04
					Within	0.06	0.37	0.15	0.22	0.03	0.04
			Polytomous	One	Between	0.00	0.31	0.16	0.16	0.02	0.04
					Within	0.07	0.37	0.15	0.23	0.03	0.04
				Separate	Between	-0.13	0.10	0.14	-0.03	0.01	0.04
					Within	-0.04	0.19	0.13	0.07	0.01	0.04
				Composite	Between	-0.14	0.12	0.18	-0.02	0.00	0.03
					Within	-0.11	0.25	0.28	0.03	0.01	0.05
			Dichotomous	One	Between	-0.14	0.12	0.18	-0.02	0.00	0.03
					Within	-0.12	0.25	0.28	0.03	0.01	0.05
				Separate	Between	-0.15	0.13	0.19	-0.02	0.01	0.03
		1			Within	-0.10	0.25	0.28	0.03	0.01	0.05
				Composite	Between	-0.15	0.06	0.13	-0.03	0.01	0.03
					Within	-0.06	0.18	0.13	0.06	0.02	0.03
			Polytomous	One	Between	-0.15	0.06	0.13	-0.03	0.01	0.03
					Within	-0.06	0.18	0.13	0.06	0.02	0.03
				Separate	Between	-0.16	0.09	0.13	-0.03	0.01	0.03
					Within	-0.04	0.22	0.13	0.07	0.02	0.03
	None			Composite	Between	-0.12	0.33	0.25	0.04	0.03	0.04
					Within	-0.11	0.28	0.24	0.04	0.03	0.04
			Dichotomous	One	Between	-0.05	0.38	0.24	0.13	0.04	0.04
					Within	-0.03	0.39	0.24	0.13	0.04	0.04
				Separate	Between	-0.13	0.13	0.19	0.01	0.00	0.03
		0.2			Within	-0.15	0.13	0.19	0.01	0.00	0.03
				Composite	Between	-0.06	0.62	0.18	0.45	0.08	0.05
					Within	-0.06	0.60	0.17	0.44	0.08	0.05
			Polytomous	One	Between	0.33	0.64	0.16	0.52	0.03	0.04
					Within	0.36	0.64	0.16	0.52	0.03	0.04
				Separate	Between	-0.11	0.11	0.15	0.01	0.01	0.04
					Within	-0.10	0.11	0.12	0.01	0.01	0.03
				Composite	Between	-0.10	0.26	0.22	0.06	0.02	0.04
					Within	-0.09	0.25	0.22	0.06	0.02	0.04
			Dichotomous	One	Between	-0.07	0.25	0.21	0.09	0.01	0.04
					Within	-0.04	0.25	0.21	0.09	0.01	0.04
				Separate	Between	-0.15	0.13	0.19	0.01	0.00	0.03
		0.5			Within	-0.16	0.13	0.20	0.01	0.00	0.03
				Composite	Between	0.01	0.52	0.18	0.27	0.06	0.05
					Within	0.05	0.48	0.19	0.27	0.05	0.05
			Polytomous	One	Between	0.11	0.49	0.18	0.31	0.03	0.05
					Within	0.13	0.46	0.18	0.31	0.03	0.05
				Separate	Between	-0.09	0.14	0.14	0.01	0.01	0.03
					Within	-0.10	0.12	0.13	0.01	0.01	0.04
				Composite	Between	-0.08	0.18	0.19	0.04	0.01	0.03
		0.7			Within	-0.09	0.18	0.19	0.04	0.01	0.03
			Dichotomous	One	Between	-0.08	0.19	0.20	0.05	0.01	0.03
					Within	-0.08	0.20	0.19	0.05	0.01	0.03
				Separate	Between	-0.12	0.13	0.19	0.01	0.00	0.03
					Within	-0.15	0.13	0.19	0.00	0.00	0.03
				Composite	Between	0.01	0.29	0.15	0.14	0.02	0.04
			Polytomous		Within	0.00	0.33	0.15	0.14	0.03	0.04
				One	Between	0.02	0.30	0.15	0.16	0.02	0.04

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
					Within	0.03	0.35	0.15	0.17	0.02	0.04
				Separate	Between	-0.09	0.12	0.13	0.01	0.01	0.03
					Within	-0.10	0.12	0.13	0.01	0.01	0.03
			Dichotomous	Composite	Between	-0.14	0.12	0.19	0.00	0.00	0.03
					Within	-0.12	0.13	0.18	0.00	0.00	0.03
				One	Between	-0.14	0.12	0.19	0.00	0.00	0.03
					Within	-0.12	0.12	0.18	0.00	0.00	0.03
				Separate	Between	-0.14	0.12	0.19	0.00	0.00	0.03
					Within	-0.14	0.13	0.19	0.01	0.00	0.03
		1	Polytomous	Composite	Between	-0.12	0.07	0.12	-0.02	0.01	0.03
					Within	-0.12	0.07	0.13	-0.02	0.01	0.03
				One	Between	-0.12	0.07	0.12	-0.02	0.01	0.03
					Within	-0.12	0.07	0.13	-0.02	0.01	0.03
				Separate	Between	-0.10	0.11	0.13	0.01	0.00	0.03
					Within	-0.10	0.10	0.13	0.01	0.01	0.03

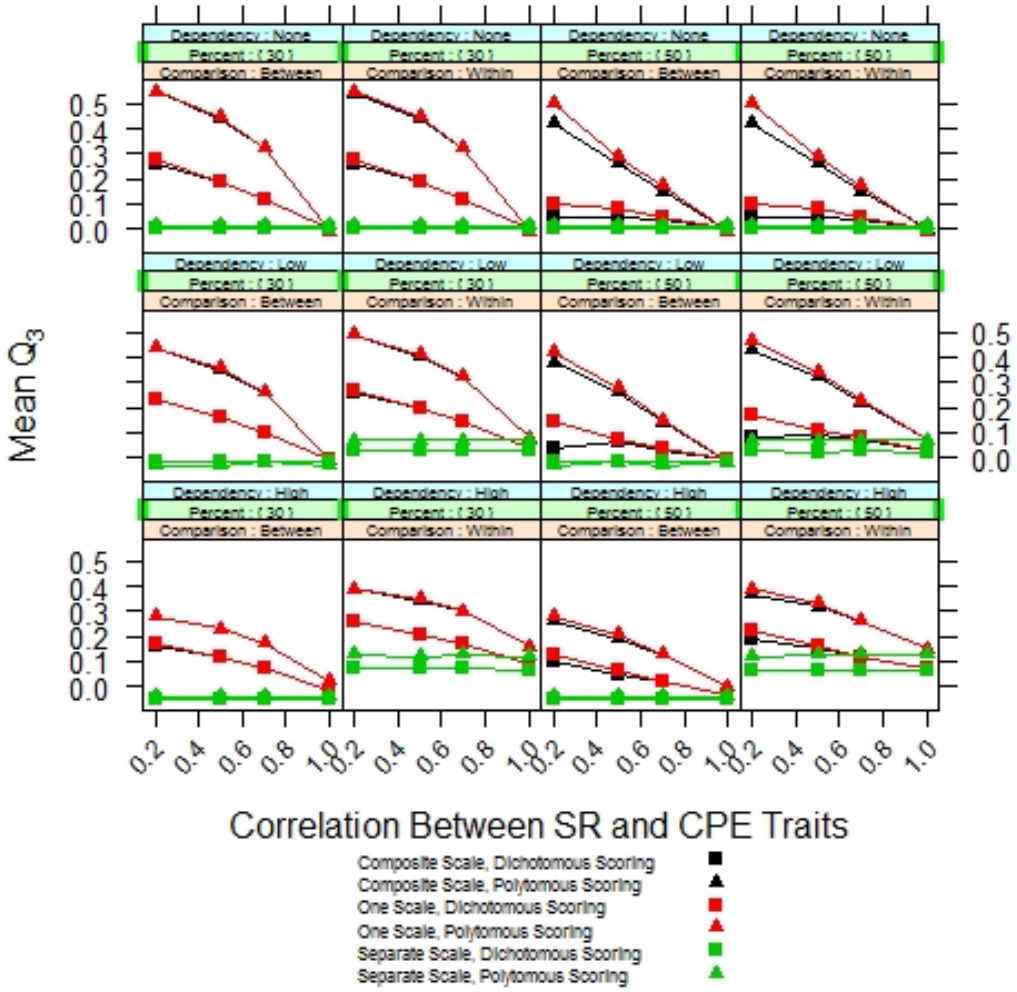

Figure 30: Average Q3 for Sample Size of 1000 and 120 Items

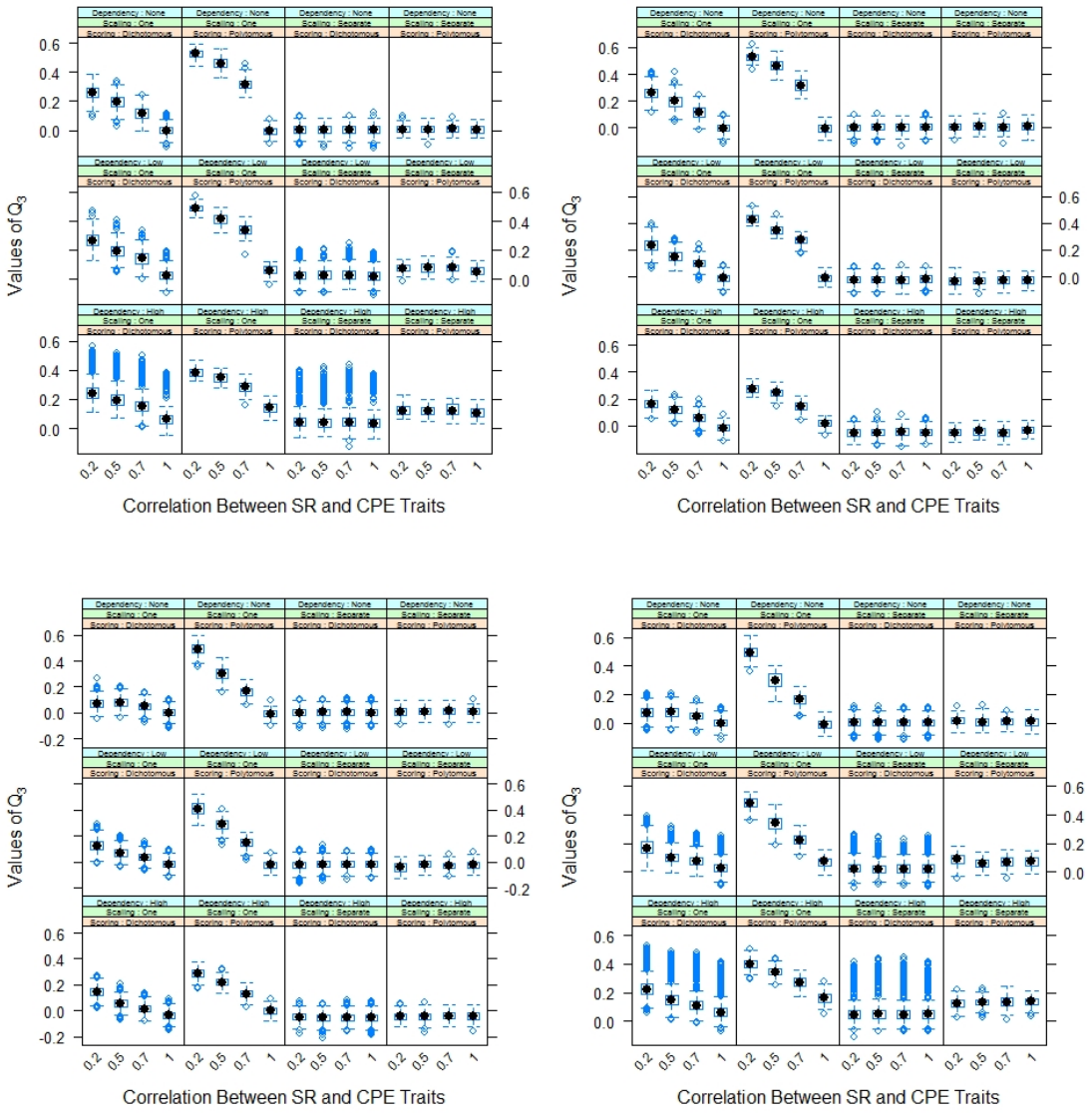

Figure 31: A-D 1000 and 120 Items for Within and Between Comparisons for 30% and 50% CPEs

Note: Clockwise from Top Left: 1) Within, 30%, 2) Between, 50%, 3) Within, 50%, 4) Between, 50%

Table 28: Simulation Results for 3,000 Sample Size and 60 Items

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
30				Composite	Between	0.01	0.32	0.23	0.16	0.01	0.04
					Within	0.06	0.58	0.42	0.26	0.01	0.08
			Dichotomous	One	Between	0.02	0.33	0.24	0.17	0.01	0.04
					Within	0.07	0.58	0.42	0.26	0.01	0.08
				Separate	Between	-0.17	0.09	0.20	-0.05	0.01	0.03
					Within	-0.10	0.46	0.47	0.07	0.01	0.09
		0.2		Composite	Between	0.16	0.40	0.15	0.28	0.01	0.04
					Within	0.28	0.54	0.14	0.39	0.01	0.04
			Polytomous	One	Between	0.16	0.40	0.15	0.28	0.01	0.04
					Within	0.29	0.54	0.14	0.39	0.01	0.03
				Separate	Between	-0.17	0.08	0.14	-0.04	0.01	0.03
					Within	-0.01	0.27	0.15	0.13	0.01	0.04
				Composite	Between	-0.02	0.28	0.22	0.12	0.01	0.04
					Within	0.03	0.55	0.43	0.21	0.01	0.08
			Dichotomous	One	Between	-0.02	0.28	0.22	0.12	0.01	0.04
					Within	0.03	0.55	0.43	0.21	0.01	0.08
				Separate	Between	-0.20	0.10	0.20	-0.05	0.01	0.03
					Within	-0.13	0.46	0.47	0.07	0.01	0.09
		0.5		Composite	Between	0.08	0.35	0.14	0.23	0.01	0.03
					Within	0.21	0.45	0.14	0.34	0.01	0.03
			Polytomous	One	Between	0.09	0.35	0.14	0.23	0.02	0.03
					Within	0.21	0.45	0.14	0.35	0.01	0.03
				Separate	Between	-0.15	0.09	0.14	-0.04	0.01	0.03
					Within	-0.01	0.27	0.16	0.12	0.01	0.04
	High			Composite	Between	-0.06	0.23	0.20	0.07	0.01	0.03
					Within	0.01	0.52	0.43	0.17	0.01	0.08
			Dichotomous	One	Between	-0.06	0.23	0.20	0.07	0.01	0.03
					Within	0.01	0.52	0.43	0.17	0.01	0.08
				Separate	Between	-0.17	0.08	0.20	-0.05	0.01	0.03
					Within	-0.12	0.44	0.47	0.07	0.01	0.09
		0.7		Composite	Between	0.04	0.30	0.15	0.17	0.01	0.04
					Within	0.15	0.40	0.15	0.30	0.01	0.04
			Polytomous	One	Between	0.05	0.30	0.15	0.17	0.01	0.04
					Within	0.16	0.42	0.15	0.30	0.02	0.04
				Separate	Between	-0.15	0.06	0.14	-0.04	0.01	0.03
					Within	-0.01	0.27	0.16	0.13	0.01	0.04
				Composite	Between	-0.15	0.11	0.19	-0.02	0.01	0.03
					Within	-0.08	0.47	0.46	0.09	0.01	0.09
			Dichotomous	One	Between	-0.15	0.11	0.19	-0.02	0.01	0.03
					Within	-0.08	0.47	0.46	0.09	0.01	0.09
				Separate	Between	-0.22	0.09	0.19	-0.05	0.01	0.03
					Within	-0.11	0.46	0.47	0.06	0.01	0.09
		1		Composite	Between	-0.10	0.16	0.14	0.02	0.01	0.03
					Within	0.01	0.29	0.15	0.16	0.02	0.04
			Polytomous	One	Between	-0.10	0.16	0.14	0.02	0.01	0.03
					Within	0.01	0.29	0.15	0.16	0.02	0.04
				Separate	Between	-0.16	0.08	0.14	-0.04	0.01	0.03
					Within	-0.02	0.28	0.16	0.12	0.02	0.04
	Low			Composite	Between	0.02	0.43	0.28	0.23	0.01	0.05
					Within	0.04	0.49	0.32	0.26	0.01	0.06
			Dichotomous	One	Between	0.04	0.44	0.29	0.23	0.01	0.05
					Within	0.05	0.50	0.32	0.27	0.01	0.06
				Separate	Between	-0.15	0.11	0.20	-0.02	0.00	0.03
					Within	-0.12	0.25	0.29	0.03	0.01	0.05
		0.2		Composite	Between	0.31	0.57	0.14	0.44	0.02	0.04
					Within	0.36	0.60	0.14	0.49	0.01	0.03
			Polytomous	One	Between	0.32	0.57	0.14	0.44	0.02	0.03
					Within	0.37	0.60	0.14	0.49	0.01	0.03
				Separate	Between	-0.16	0.09	0.14	-0.03	0.01	0.03
					Within	-0.07	0.18	0.15	0.07	0.01	0.04
				Composite	Between	-0.01	0.33	0.25	0.16	0.01	0.04
					Within	0.02	0.42	0.31	0.20	0.01	0.05
		0.5	Dichotomous	One	Between	-0.01	0.33	0.25	0.16	0.01	0.04
					Within	0.02	0.42	0.31	0.20	0.01	0.05
				Separate	Between	-0.16	0.10	0.20	-0.02	0.01	0.03
					Within	-0.12	0.24	0.29	0.03	0.01	0.05

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
				Composite	Between	0.21	0.50	0.16	0.35	0.02	0.04
					Within	0.26	0.56	0.16	0.40	0.02	0.04
			Polytomous	One	Between	0.22	0.51	0.16	0.36	0.02	0.04
					Within	0.25	0.56	0.16	0.41	0.02	0.04
				Separate	Between	-0.16	0.11	0.15	-0.03	0.01	0.03
					Within	-0.03	0.20	0.15	0.07	0.02	0.04
				Composite	Between	-0.04	0.26	0.22	0.10	0.01	0.04
					Within	-0.02	0.39	0.29	0.14	0.01	0.05
			Dichotomous	One	Between	-0.04	0.27	0.22	0.10	0.01	0.04
					Within	-0.01	0.40	0.29	0.14	0.01	0.05
				Separate	Between	-0.17	0.12	0.20	-0.02	0.01	0.03
					Within	-0.11	0.27	0.29	0.03	0.01	0.05
		0.7		Composite	Between	0.11	0.42	0.18	0.26	0.02	0.04
					Within	0.17	0.46	0.17	0.32	0.02	0.04
			Polytomous	One	Between	0.12	0.42	0.17	0.26	0.02	0.04
					Within	0.17	0.46	0.17	0.32	0.02	0.04
				Separate	Between	-0.13	0.09	0.15	-0.02	0.01	0.03
					Within	-0.06	0.21	0.15	0.07	0.02	0.04
				Composite	Between	-0.14	0.12	0.19	-0.01	0.00	0.03
					Within	-0.12	0.28	0.30	0.04	0.01	0.05
			Dichotomous	One	Between	-0.14	0.12	0.19	-0.01	0.00	0.03
					Within	-0.12	0.28	0.30	0.04	0.01	0.05
				Separate	Between	-0.16	0.12	0.20	-0.02	0.00	0.03
					Within	-0.12	0.27	0.29	0.03	0.01	0.05
		1		Composite	Between	-0.11	0.09	0.14	-0.01	0.01	0.03
					Within	-0.06	0.21	0.16	0.08	0.02	0.04
			Polytomous	One	Between	-0.12	0.09	0.14	-0.01	0.01	0.03
					Within	-0.06	0.21	0.16	0.08	0.02	0.04
				Separate	Between	-0.13	0.09	0.14	-0.03	0.01	0.03
					Within	-0.06	0.19	0.14	0.07	0.01	0.04
	None			Composite	Between	0.06	0.47	0.31	0.26	0.01	0.06
					Within	0.07	0.47	0.31	0.26	0.01	0.06
			Dichotomous	One	Between	0.06	0.50	0.32	0.28	0.01	0.06
					Within	0.07	0.48	0.31	0.28	0.01	0.06
				Separate	Between	-0.13	0.18	0.20	0.00	0.00	0.03
					Within	-0.14	0.14	0.20	0.00	0.00	0.03
		0.2		Composite	Between	0.41	0.68	0.16	0.55	0.02	0.04
					Within	0.37	0.70	0.16	0.54	0.02	0.04
			Polytomous	One	Between	0.42	0.69	0.16	0.55	0.02	0.04
					Within	0.37	0.70	0.16	0.55	0.02	0.04
				Separate	Between	-0.10	0.11	0.15	0.01	0.00	0.03
					Within	-0.09	0.11	0.14	0.01	0.00	0.03
				Composite	Between	0.01	0.43	0.29	0.19	0.01	0.05
					Within	0.01	0.40	0.28	0.19	0.01	0.05
			Dichotomous	One	Between	0.01	0.42	0.29	0.19	0.01	0.05
					Within	0.00	0.40	0.28	0.19	0.01	0.05
				Separate	Between	-0.12	0.14	0.20	0.00	0.00	0.03
					Within	-0.14	0.12	0.20	0.00	0.00	0.03
		0.5		Composite	Between	0.19	0.59	0.18	0.44	0.02	0.04
					Within	0.26	0.58	0.18	0.44	0.02	0.04
			Polytomous	One	Between	0.19	0.59	0.18	0.45	0.02	0.04
					Within	0.28	0.59	0.18	0.45	0.02	0.04
				Separate	Between	-0.12	0.11	0.15	0.01	0.01	0.03
					Within	-0.11	0.11	0.14	0.01	0.01	0.03
				Composite	Between	-0.04	0.28	0.24	0.12	0.01	0.04
					Within	-0.04	0.28	0.24	0.12	0.01	0.04
			Dichotomous	One	Between	-0.04	0.28	0.24	0.12	0.01	0.04
					Within	-0.04	0.28	0.24	0.12	0.01	0.04
				Separate	Between	-0.14	0.13	0.20	0.00	0.00	0.03
		0.7			Within	-0.14	0.13	0.20	0.00	0.00	0.03
				Composite	Between	0.15	0.48	0.19	0.32	0.02	0.05
					Within	0.12	0.46	0.18	0.32	0.02	0.05
			Polytomous	One	Between	0.15	0.48	0.19	0.32	0.02	0.05
					Within	0.13	0.47	0.18	0.32	0.02	0.05
				Separate	Between	-0.12	0.14	0.14	0.01	0.00	0.03
					Within	-0.10	0.11	0.14	0.01	0.01	0.03
		1	Dichotomous	Composite	Between	-0.13	0.13	0.20	0.00	0.00	0.03
					Within	-0.16	0.13	0.20	0.00	0.00	0.03
				One	Between	-0.13	0.13	0.20	0.00	0.00	0.03

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
					Within	-0.16	0.13	0.20	0.00	0.00	0.03
				Separate	Between	-0.13	0.13	0.20	0.00	0.00	0.03
					Within	-0.15	0.14	0.20	0.00	0.00	0.03
			Polytomous	Composite	Between	-0.12	0.10	0.14	-0.01	0.01	0.03
					Within	-0.13	0.10	0.14	-0.01	0.01	0.03
				One	Between	-0.12	0.10	0.14	-0.01	0.01	0.03
					Within	-0.13	0.10	0.14	-0.01	0.01	0.03
				Separate	Between	-0.10	0.11	0.14	0.01	0.01	0.03
					Within	-0.11	0.11	0.14	0.01	0.01	0.03
50											
			Dichotomous	Composite	Between	-0.10	0.31	0.25	0.10	0.03	0.04
					Within	-0.06	0.55	0.46	0.19	0.03	0.07
				One	Between	-0.04	0.33	0.25	0.13	0.01	0.04
					Within	-0.04	0.56	0.45	0.22	0.01	0.07
				Separate	Between	-0.18	0.09	0.21	-0.05	0.01	0.03
		0.2			Within	-0.14	0.46	0.48	0.06	0.01	0.07
			Polytomous	Composite	Between	0.12	0.39	0.18	0.26	0.02	0.04
					Within	0.23	0.50	0.17	0.37	0.02	0.04
				One	Between	0.14	0.40	0.17	0.28	0.02	0.04
					Within	0.25	0.51	0.17	0.39	0.01	0.04
				Separate	Between	-0.16	0.08	0.17	-0.04	0.01	0.03
					Within	-0.01	0.28	0.18	0.12	0.01	0.04
			Dichotomous	Composite	Between	-0.11	0.23	0.23	0.05	0.01	0.04
					Within	-0.02	0.51	0.45	0.15	0.01	0.07
				One	Between	-0.08	0.22	0.22	0.06	0.01	0.03
					Within	-0.01	0.51	0.45	0.16	0.01	0.07
				Separate	Between	-0.21	0.08	0.22	-0.05	0.01	0.03
		0.5			Within	-0.09	0.45	0.48	0.06	0.01	0.07
			Polytomous	Composite	Between	0.04	0.33	0.18	0.19	0.02	0.04
					Within	0.16	0.45	0.17	0.32	0.01	0.04
				One	Between	0.06	0.33	0.17	0.21	0.02	0.04
					Within	0.17	0.45	0.17	0.33	0.01	0.04
				Separate	Between	-0.17	0.07	0.16	-0.04	0.01	0.03
	High				Within	-0.02	0.28	0.18	0.13	0.01	0.04
			Dichotomous	Composite	Between	-0.12	0.16	0.21	0.02	0.01	0.03
					Within	-0.06	0.49	0.46	0.12	0.01	0.07
				One	Between	-0.12	0.16	0.21	0.02	0.01	0.03
					Within	-0.05	0.49	0.46	0.12	0.01	0.07
				Separate	Between	-0.18	0.09	0.22	-0.05	0.01	0.03
		0.7			Within	-0.13	0.47	0.49	0.06	0.01	0.08
			Polytomous	Composite	Between	-0.01	0.27	0.18	0.13	0.01	0.04
					Within	0.12	0.41	0.18	0.26	0.02	0.04
				One	Between	-0.01	0.27	0.18	0.13	0.01	0.04
					Within	0.13	0.41	0.18	0.26	0.02	0.04
				Separate	Between	-0.16	0.09	0.17	-0.04	0.01	0.03
					Within	0.00	0.27	0.18	0.13	0.01	0.04
			Dichotomous	Composite	Between	-0.16	0.11	0.21	-0.03	0.00	0.03
					Within	-0.11	0.49	0.48	0.07	0.01	0.07
				One	Between	-0.16	0.11	0.21	-0.03	0.00	0.03
					Within	-0.11	0.49	0.49	0.07	0.01	0.07
				Separate	Between	-0.18	0.10	0.22	-0.05	0.01	0.03
		1			Within	-0.11	0.47	0.49	0.06	0.01	0.07
			Polytomous	Composite	Between	-0.13	0.12	0.17	0.00	0.01	0.03
					Within	0.00	0.31	0.18	0.15	0.02	0.04
				One	Between	-0.13	0.12	0.17	0.00	0.01	0.03
					Within	-0.01	0.31	0.18	0.15	0.02	0.04
				Separate	Between	-0.17	0.09	0.17	-0.04	0.01	0.03
					Within	-0.03	0.27	0.18	0.13	0.01	0.04
	Low										
			Dichotomous	Composite	Between	-0.15	0.28	0.28	0.04	0.02	0.04
					Within	-0.08	0.37	0.33	0.08	0.02	0.05
				One	Between	-0.05	0.38	0.28	0.14	0.04	0.04
					Within	-0.06	0.43	0.33	0.17	0.04	0.05
				Separate	Between	-0.18	0.12	0.22	-0.02	0.01	0.03
		0.2			Within	-0.12	0.31	0.31	0.03	0.01	0.04
			Polytomous	Composite	Between	0.19	0.53	0.20	0.38	0.02	0.04
					Within	0.24	0.58	0.21	0.43	0.02	0.04
				One	Between	0.25	0.57	0.19	0.42	0.02	0.04
					Within	0.32	0.60	0.19	0.47	0.02	0.04
				Separate	Between	-0.15	0.10	0.17	-0.03	0.01	0.03
					Within	-0.06	0.21	0.17	0.07	0.02	0.03

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
				Composite	Between	-0.11	0.25	0.24	0.06	0.02	0.04
					Within	-0.06	0.33	0.31	0.09	0.02	0.04
			Dichotomous	One	Between	-0.07	0.23	0.24	0.07	0.01	0.04
					Within	-0.04	0.35	0.31	0.11	0.01	0.04
				Separate	Between	-0.15	0.13	0.22	-0.02	0.00	0.03
		0.5			Within	-0.12	0.27	0.31	0.02	0.00	0.04
				Composite	Between	0.06	0.43	0.21	0.26	0.02	0.04
					Within	0.14	0.48	0.21	0.32	0.02	0.04
			Polytomous	One	Between	0.12	0.45	0.20	0.28	0.02	0.04
					Within	0.15	0.48	0.21	0.34	0.02	0.04
				Separate	Between	-0.13	0.10	0.16	-0.02	0.01	0.03
					Within	-0.09	0.20	0.18	0.07	0.01	0.03
				Composite	Between	-0.11	0.21	0.22	0.03	0.01	0.03
					Within	-0.07	0.31	0.31	0.07	0.01	0.04
			Dichotomous	One	Between	-0.10	0.22	0.22	0.04	0.01	0.03
					Within	-0.06	0.31	0.31	0.08	0.01	0.04
				Separate	Between	-0.15	0.13	0.22	-0.02	0.01	0.03
		0.7			Within	-0.12	0.27	0.31	0.03	0.01	0.04
				Composite	Between	0.00	0.28	0.20	0.14	0.02	0.04
					Within	0.07	0.38	0.20	0.22	0.02	0.04
			Polytomous	One	Between	0.01	0.29	0.20	0.15	0.01	0.04
					Within	0.07	0.38	0.20	0.23	0.02	0.04
				Separate	Between	-0.16	0.10	0.16	-0.03	0.01	0.03
					Within	-0.06	0.22	0.17	0.07	0.01	0.04
				Composite	Between	-0.14	0.14	0.21	-0.01	0.00	0.03
					Within	-0.11	0.27	0.31	0.03	0.01	0.04
			Dichotomous	One	Between	-0.14	0.14	0.21	-0.01	0.00	0.03
					Within	-0.11	0.27	0.31	0.03	0.01	0.04
				Separate	Between	-0.17	0.14	0.22	-0.02	0.00	0.03
		1			Within	-0.11	0.27	0.31	0.02	0.00	0.04
				Composite	Between	-0.14	0.09	0.16	-0.02	0.01	0.03
					Within	-0.06	0.21	0.18	0.07	0.01	0.04
			Polytomous	One	Between	-0.14	0.09	0.16	-0.02	0.01	0.03
					Within	-0.06	0.21	0.18	0.07	0.01	0.04
				Separate	Between	-0.14	0.09	0.16	-0.02	0.01	0.03
					Within	-0.07	0.21	0.17	0.07	0.01	0.03
	None			Composite	Between	-0.13	0.30	0.29	0.05	0.02	0.04
					Within	-0.13	0.30	0.29	0.05	0.02	0.04
			Dichotomous	One	Between	-0.07	0.40	0.27	0.10	0.04	0.04
					Within	-0.08	0.45	0.26	0.10	0.04	0.04
				Separate	Between	-0.14	0.14	0.21	0.00	0.00	0.03
		0.2			Within	-0.15	0.16	0.22	0.00	0.00	0.03
				Composite	Between	0.01	0.63	0.24	0.42	0.07	0.05
					Within	0.01	0.63	0.23	0.42	0.07	0.05
			Polytomous	One	Between	0.27	0.67	0.22	0.50	0.02	0.04
					Within	0.30	0.67	0.21	0.50	0.02	0.04
				Separate	Between	-0.11	0.11	0.17	0.01	0.00	0.03
					Within	-0.13	0.14	0.17	0.01	0.00	0.03
				Composite	Between	-0.09	0.29	0.25	0.05	0.01	0.04
					Within	-0.09	0.25	0.24	0.05	0.01	0.04
			Dichotomous	One	Between	-0.06	0.26	0.24	0.08	0.01	0.04
					Within	-0.06	0.27	0.24	0.08	0.01	0.04
				Separate	Between	-0.14	0.16	0.22	0.00	0.00	0.03
		0.5			Within	-0.13	0.13	0.22	0.00	0.00	0.03
				Composite	Between	0.00	0.47	0.23	0.26	0.04	0.05
					Within	0.00	0.51	0.22	0.26	0.04	0.05
			Polytomous	One	Between	0.09	0.49	0.23	0.29	0.03	0.05
					Within	0.08	0.50	0.22	0.29	0.03	0.05
				Separate	Between	-0.11	0.16	0.16	0.01	0.00	0.03
					Within	-0.11	0.13	0.16	0.01	0.00	0.03
				Composite	Between	-0.10	0.22	0.23	0.04	0.00	0.03
		0.7			Within	-0.09	0.19	0.22	0.04	0.00	0.03
			Dichotomous	One	Between	-0.09	0.23	0.23	0.05	0.00	0.04
					Within	-0.08	0.20	0.22	0.05	0.00	0.04
				Separate	Between	-0.15	0.18	0.22	0.00	0.00	0.03
					Within	-0.14	0.14	0.22	0.00	0.00	0.03
			Polytomous	Composite	Between	0.00	0.30	0.20	0.15	0.02	0.04
					Within	0.00	0.31	0.20	0.15	0.02	0.04
				One	Between	0.02	0.32	0.20	0.17	0.02	0.04

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
					Within	0.02	0.32	0.20	0.17	0.02	0.04
				Separate	Between	-0.11	0.12	0.17	0.01	0.00	0.03
					Within	-0.10	0.14	0.16	0.01	0.00	0.03
			Dichotomous	Composite	Between	-0.13	0.13	0.21	0.00	0.00	0.03
					Within	-0.14	0.17	0.21	0.00	0.00	0.03
				One	Between	-0.13	0.13	0.21	0.00	0.00	0.03
					Within	-0.14	0.17	0.21	0.00	0.00	0.03
				Separate	Between	-0.12	0.13	0.21	0.00	0.00	0.03
					Within	-0.14	0.18	0.21	0.00	0.00	0.03
		1		Composite	Between	-0.11	0.10	0.16	-0.01	0.00	0.03
					Within	-0.13	0.11	0.17	0.00	0.00	0.03
			Polytomous	One	Between	-0.11	0.10	0.16	-0.01	0.00	0.03
					Within	-0.13	0.11	0.17	-0.01	0.00	0.03
				Separate	Between	-0.10	0.12	0.16	0.01	0.00	0.03
					Within	-0.12	0.13	0.17	0.01	0.00	0.03

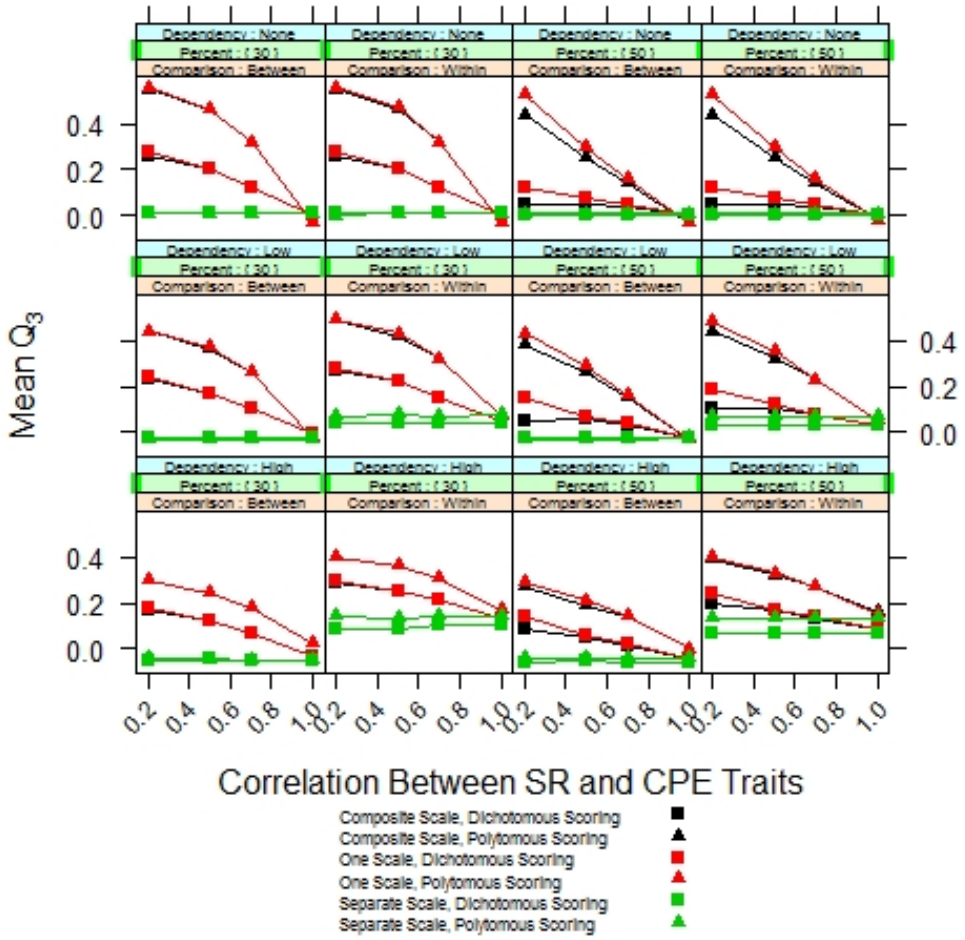

Figure 32: Average Q3 for Sample Size of 3000 and 60 Items

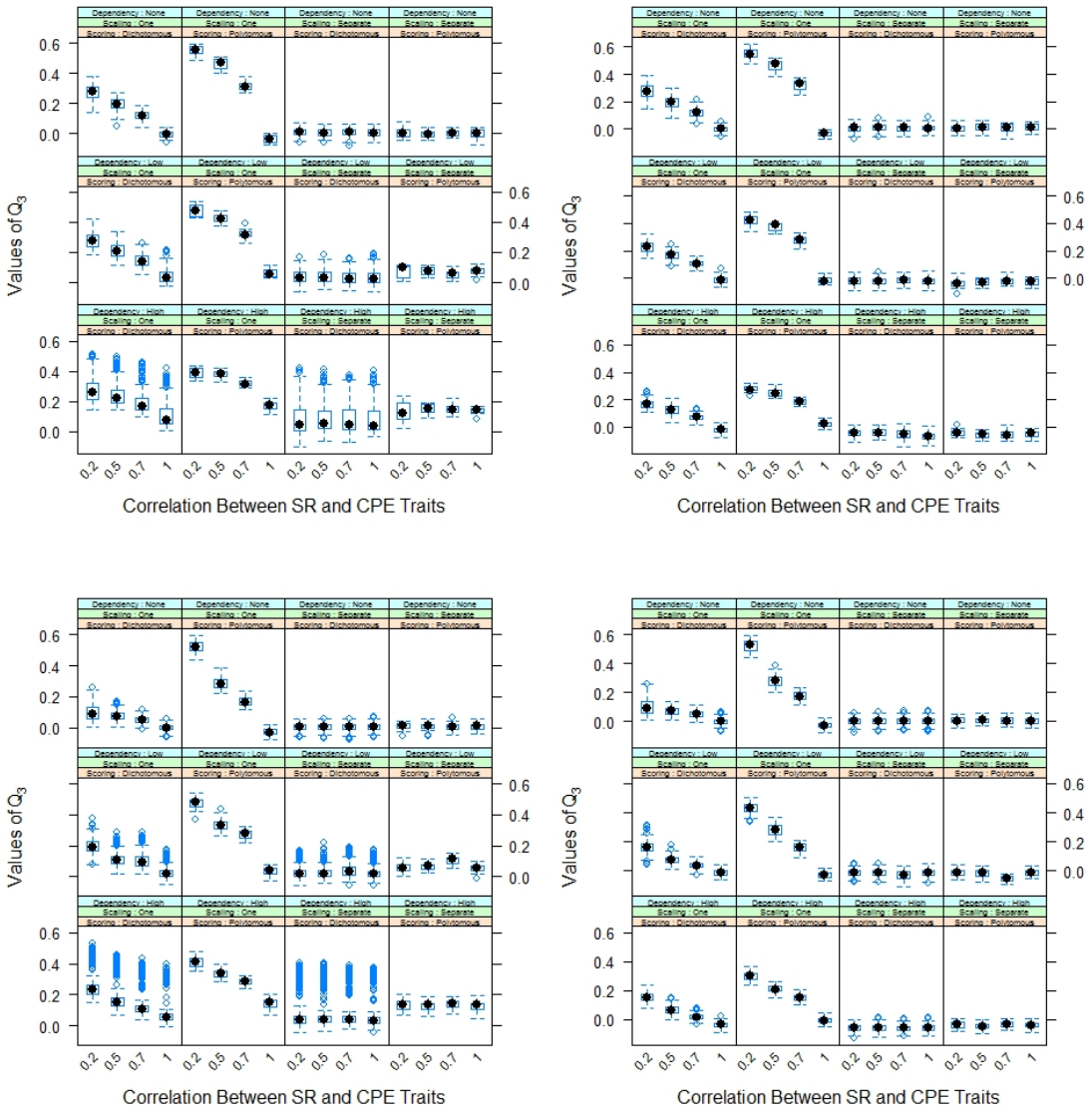

Figure 33: A-D 3000 and 60 Items for Within and Between Comparisons for 30% and 50% CPEs

Note: Clockwise from Top Left: 1) Within, 30%, 2) Between, 50%, 3) Within, 50%, 4) Between, 50%

Table 29: Simulation Results for 3,000 Sample Size and 120 Items

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
30					Between	0.06	0.32	0.18	0.16	0.01	0.03
				Composite	Within	0.09	0.58	0.39	0.26	0.01	0.08
					Between	0.06	0.33	0.18	0.17	0.01	0.03
			Dichotomous	One	Within	0.09	0.59	0.39	0.26	0.01	0.08
					Between	-0.14	0.03	0.13	-0.05	0.01	0.02
				Separate	Within	-0.08	0.42	0.42	0.06	0.01	0.09
		0.2			Between	0.19	0.38	0.11	0.28	0.01	0.03
				Composite	Within	0.29	0.50	0.11	0.39	0.01	0.03
					Between	0.19	0.39	0.11	0.28	0.01	0.03
			Polytomous	One	Within	0.29	0.50	0.11	0.39	0.01	0.03
					Between	-0.12	0.02	0.09	-0.04	0.01	0.02
				Separate	Within	0.03	0.23	0.11	0.13	0.01	0.03
					Between	0.03	0.24	0.16	0.12	0.01	0.03
				Composite	Within	0.08	0.52	0.38	0.22	0.01	0.08
					Between	0.03	0.24	0.16	0.12	0.01	0.03
			Dichotomous	One	Within	0.09	0.52	0.38	0.22	0.01	0.08
					Between	-0.16	0.03	0.13	-0.06	0.01	0.02
				Separate	Within	-0.06	0.43	0.42	0.07	0.01	0.09
		0.5			Between	0.15	0.33	0.11	0.23	0.01	0.03
				Composite	Within	0.25	0.44	0.10	0.35	0.02	0.03
					Between	0.15	0.34	0.11	0.24	0.01	0.03
			Polytomous	One	Within	0.25	0.44	0.11	0.36	0.02	0.03
					Between	-0.12	0.03	0.09	-0.04	0.01	0.02
				Separate	Within	0.03	0.25	0.11	0.13	0.02	0.03
	High				Between	-0.02	0.16	0.13	0.07	0.01	0.02
				Composite	Within	0.03	0.51	0.39	0.17	0.01	0.08
					Between	-0.02	0.16	0.13	0.07	0.01	0.02
			Dichotomous	One	Within	0.03	0.51	0.39	0.17	0.01	0.08
					Between	-0.14	0.04	0.13	-0.06	0.01	0.02
				Separate	Within	-0.06	0.45	0.42	0.07	0.01	0.09
		0.7			Between	0.06	0.28	0.11	0.17	0.01	0.03
				Composite	Within	0.19	0.40	0.11	0.30	0.02	0.03
					Between	0.06	0.28	0.11	0.17	0.01	0.03
			Polytomous	One	Within	0.19	0.40	0.11	0.30	0.02	0.03
					Between	-0.12	0.03	0.09	-0.04	0.01	0.02
				Separate	Within	0.01	0.22	0.11	0.13	0.01	0.03
					Between	-0.09	0.06	0.11	-0.02	0.00	0.02
				Composite	Within	-0.03	0.46	0.40	0.09	0.01	0.09
					Between	-0.09	0.06	0.11	-0.02	0.00	0.02
			Dichotomous	One	Within	-0.03	0.46	0.40	0.09	0.01	0.09
					Between	-0.15	0.04	0.13	-0.05	0.01	0.02
				Separate	Within	-0.05	0.43	0.42	0.06	0.01	0.09
		1			Between	-0.05	0.09	0.09	0.02	0.01	0.02
				Composite	Within	0.04	0.27	0.12	0.16	0.02	0.03
					Between	-0.05	0.09	0.09	0.02	0.01	0.02
			Polytomous	One	Within	0.04	0.28	0.12	0.16	0.02	0.03
					Between	-0.12	0.04	0.09	-0.04	0.01	0.02
				Separate	Within	0.03	0.22	0.10	0.12	0.01	0.03
					Between	0.05	0.39	0.24	0.22	0.01	0.04
				Composite	Within	0.05	0.48	0.30	0.25	0.01	0.05
					Between	0.05	0.41	0.25	0.23	0.01	0.04
			Dichotomous	One	Within	0.06	0.49	0.30	0.27	0.01	0.05
					Between	-0.13	0.07	0.13	-0.02	0.01	0.02
				Separate	Within	-0.08	0.23	0.24	0.03	0.01	0.04
		0.2			Between	0.30	0.54	0.14	0.43	0.02	0.03
				Composite	Within	0.36	0.60	0.14	0.48	0.02	0.03
					Between	0.31	0.55	0.14	0.44	0.02	0.03
			Polytomous	One	Within	0.37	0.60	0.14	0.49	0.02	0.03
					Between	-0.11	0.05	0.09	-0.03	0.01	0.02
				Separate	Within	-0.02	0.18	0.10	0.07	0.02	0.02
					Between	0.03	0.35	0.21	0.16	0.01	0.04
				Composite	Within	0.03	0.43	0.28	0.20	0.01	0.05
					Between	0.03	0.35	0.21	0.16	0.01	0.04
			Dichotomous	One	Within	0.03	0.42	0.28	0.20	0.01	0.05
					Between	-0.11	0.06	0.13	-0.02	0.00	0.02
				Separate	Within	-0.08	0.25	0.23	0.03	0.01	0.04
		0.5			Between	0.21	0.48	0.14	0.35	0.02	0.04
			Polytomous	Composite	Within	0.29	0.53	0.14	0.41	0.02	0.04

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
				One	Between	0.22	0.49	0.14	0.36	0.02	0.04
					Within	0.28	0.53	0.14	0.41	0.02	0.04
				Separate	Between	-0.11	0.05	0.09	-0.03	0.01	0.02
					Within	-0.01	0.16	0.09	0.07	0.01	0.02
		0.7	Dichotomous	Composite	Between	0.00	0.22	0.16	0.10	0.00	0.03
					Within	-0.01	0.33	0.25	0.14	0.01	0.04
				One	Between	0.00	0.22	0.16	0.10	0.00	0.03
					Within	-0.01	0.33	0.25	0.14	0.01	0.04
				Separate	Between	-0.12	0.06	0.13	-0.02	0.00	0.02
					Within	-0.07	0.25	0.24	0.03	0.00	0.04
			Polytomous	Composite	Between	0.15	0.38	0.13	0.25	0.01	0.03
					Within	0.19	0.45	0.14	0.32	0.02	0.03
				One	Between	0.15	0.38	0.13	0.25	0.01	0.03
					Within	0.19	0.45	0.14	0.32	0.02	0.03
				Separate	Between	-0.11	0.05	0.09	-0.02	0.01	0.02
					Within	-0.03	0.16	0.10	0.07	0.01	0.02
		1	Dichotomous	Composite	Between	-0.09	0.06	0.11	-0.01	0.00	0.02
					Within	-0.06	0.25	0.24	0.04	0.01	0.04
				One	Between	-0.09	0.06	0.11	-0.01	0.00	0.02
					Within	-0.06	0.25	0.24	0.04	0.01	0.04
				Separate	Between	-0.12	0.06	0.13	-0.02	0.00	0.02
					Within	-0.09	0.24	0.24	0.03	0.00	0.04
			Polytomous	Composite	Between	-0.08	0.06	0.09	-0.01	0.01	0.02
					Within	-0.02	0.21	0.11	0.08	0.02	0.03
				One	Between	-0.08	0.06	0.09	-0.01	0.01	0.02
					Within	-0.02	0.21	0.11	0.08	0.02	0.03
				Separate	Between	-0.09	0.07	0.09	-0.03	0.01	0.02
					Within	-0.02	0.19	0.09	0.07	0.01	0.02
	None	0.2	Dichotomous	Composite	Between	0.03	0.45	0.28	0.25	0.02	0.05
					Within	0.03	0.41	0.28	0.25	0.01	0.05
				One	Between	0.03	0.48	0.29	0.27	0.01	0.05
					Within	0.04	0.44	0.29	0.27	0.01	0.05
				Separate	Between	-0.08	0.07	0.12	0.00	0.00	0.02
					Within	-0.08	0.08	0.12	0.00	0.00	0.02
			Polytomous	Composite	Between	0.29	0.64	0.14	0.54	0.02	0.03
					Within	0.37	0.65	0.14	0.54	0.02	0.03
				One	Between	0.29	0.65	0.14	0.55	0.02	0.03
					Within	0.37	0.65	0.14	0.55	0.02	0.03
				Separate	Between	-0.06	0.08	0.09	0.01	0.00	0.02
					Within	-0.06	0.08	0.09	0.01	0.00	0.02
		0.5	Dichotomous	Composite	Between	0.02	0.36	0.25	0.19	0.01	0.05
					Within	0.02	0.37	0.26	0.19	0.01	0.05
				One	Between	0.03	0.36	0.25	0.19	0.01	0.05
					Within	0.02	0.37	0.26	0.19	0.01	0.05
				Separate	Between	-0.08	0.08	0.12	0.00	0.00	0.02
					Within	-0.09	0.08	0.12	0.00	0.00	0.02
			Polytomous	Composite	Between	0.26	0.61	0.17	0.44	0.02	0.04
					Within	0.27	0.62	0.17	0.44	0.02	0.04
				One	Between	0.27	0.61	0.17	0.45	0.02	0.04
					Within	0.28	0.63	0.17	0.45	0.02	0.04
				Separate	Between	-0.06	0.08	0.09	0.01	0.00	0.02
					Within	-0.06	0.08	0.09	0.01	0.00	0.02
		0.7	Dichotomous	Composite	Between	-0.01	0.25	0.18	0.11	0.01	0.03
					Within	0.00	0.24	0.18	0.11	0.01	0.03
				One	Between	-0.01	0.25	0.18	0.12	0.01	0.03
					Within	-0.01	0.25	0.19	0.12	0.01	0.03
				Separate	Between	-0.08	0.08	0.12	0.00	0.00	0.02
					Within	-0.10	0.08	0.12	0.00	0.00	0.02
			Polytomous	Composite	Between	0.15	0.45	0.16	0.32	0.02	0.04
					Within	0.18	0.47	0.16	0.32	0.01	0.04
				One	Between	0.15	0.45	0.16	0.32	0.02	0.04
					Within	0.18	0.46	0.16	0.32	0.01	0.04
				Separate	Between	-0.06	0.08	0.09	0.01	0.00	0.02
					Within	-0.06	0.08	0.09	0.01	0.00	0.02
		1	Dichotomous	Composite	Between	-0.08	0.07	0.12	0.00	0.00	0.02
					Within	-0.09	0.08	0.11	0.00	0.00	0.02
				One	Between	-0.08	0.07	0.12	0.00	0.00	0.02
					Within	-0.09	0.08	0.11	0.00	0.00	0.02
				Separate	Between	-0.09	0.08	0.12	0.00	0.00	0.02

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
					Within	-0.10	0.08	0.12	0.00	0.00	0.02
				Composite	Between	-0.07	0.06	0.08	-0.01	0.00	0.02
					Within	-0.08	0.05	0.08	-0.01	0.00	0.02
			Polytomous	One	Between	-0.07	0.06	0.08	-0.01	0.00	0.02
					Within	-0.08	0.05	0.08	-0.01	0.00	0.02
				Separate	Between	-0.06	0.08	0.09	0.01	0.00	0.02
					Within	-0.06	0.08	0.09	0.01	0.00	0.02
50				Composite	Between	-0.05	0.28	0.19	0.10	0.02	0.03
					Within	0.03	0.52	0.41	0.19	0.01	0.07
			Dichotomous	One	Between	0.00	0.31	0.20	0.14	0.01	0.03
					Within	0.06	0.55	0.41	0.22	0.01	0.06
				Separate	Between	-0.15	0.03	0.14	-0.05	0.01	0.02
					Within	-0.04	0.44	0.42	0.06	0.01	0.07
		0.2		Composite	Between	0.15	0.36	0.14	0.25	0.01	0.03
					Within	0.26	0.48	0.14	0.37	0.01	0.03
			Polytomous	One	Between	0.17	0.38	0.14	0.28	0.01	0.03
					Within	0.29	0.51	0.14	0.39	0.01	0.03
				Separate	Between	-0.13	0.04	0.11	-0.04	0.01	0.02
					Within	0.01	0.24	0.13	0.12	0.01	0.02
				Composite	Between	-0.05	0.18	0.16	0.05	0.01	0.03
					Within	0.01	0.50	0.41	0.15	0.01	0.07
			Dichotomous	One	Between	-0.03	0.17	0.15	0.06	0.01	0.02
					Within	0.02	0.50	0.40	0.15	0.01	0.07
				Separate	Between	-0.14	0.04	0.14	-0.05	0.00	0.02
					Within	-0.05	0.45	0.43	0.06	0.00	0.07
		0.5		Composite	Between	0.07	0.29	0.14	0.19	0.01	0.03
					Within	0.20	0.42	0.14	0.31	0.01	0.03
			Polytomous	One	Between	0.10	0.30	0.14	0.20	0.01	0.03
					Within	0.21	0.45	0.14	0.33	0.01	0.03
				Separate	Between	-0.12	0.05	0.11	-0.04	0.01	0.02
					Within	0.03	0.24	0.13	0.12	0.01	0.03
	High			Composite	Between	-0.07	0.10	0.13	0.02	0.01	0.02
					Within	-0.01	0.45	0.41	0.12	0.01	0.07
			Dichotomous	One	Between	-0.07	0.10	0.13	0.02	0.01	0.02
					Within	-0.01	0.45	0.41	0.12	0.01	0.07
				Separate	Between	-0.14	0.04	0.14	-0.05	0.00	0.02
					Within	-0.05	0.42	0.42	0.06	0.01	0.07
		0.7		Composite	Between	0.01	0.22	0.13	0.13	0.01	0.03
					Within	0.15	0.36	0.14	0.26	0.01	0.03
			Polytomous	One	Between	0.02	0.22	0.12	0.13	0.01	0.03
					Within	0.15	0.36	0.13	0.26	0.01	0.03
				Separate	Between	-0.13	0.04	0.11	-0.04	0.01	0.02
					Within	0.02	0.22	0.12	0.12	0.01	0.03
				Composite	Between	-0.12	0.05	0.12	-0.03	0.00	0.02
					Within	-0.04	0.45	0.42	0.07	0.01	0.07
			Dichotomous	One	Between	-0.12	0.05	0.12	-0.03	0.00	0.02
					Within	-0.04	0.45	0.42	0.07	0.01	0.07
				Separate	Between	-0.14	0.03	0.14	-0.05	0.00	0.02
					Within	-0.06	0.44	0.42	0.06	0.00	0.07
		1		Composite	Between	-0.09	0.09	0.11	0.00	0.01	0.02
					Within	0.05	0.25	0.13	0.15	0.01	0.03
			Polytomous	One	Between	-0.09	0.08	0.11	0.00	0.01	0.02
					Within	0.05	0.25	0.13	0.15	0.01	0.03
				Separate	Between	-0.12	0.04	0.11	-0.04	0.01	0.02
					Within	0.02	0.22	0.12	0.12	0.01	0.02
	Low			Composite	Between	-0.08	0.30	0.23	0.06	0.03	0.04
					Within	-0.04	0.35	0.29	0.10	0.03	0.04
			Dichotomous	One	Between	-0.03	0.34	0.22	0.13	0.04	0.04
					Within	-0.01	0.40	0.29	0.17	0.03	0.04
				Separate	Between	-0.12	0.06	0.13	-0.02	0.00	0.02
					Within	-0.07	0.25	0.25	0.02	0.00	0.03
		0.2		Composite	Between	0.21	0.51	0.17	0.37	0.02	0.04
					Within	0.29	0.57	0.17	0.42	0.02	0.04
			Polytomous	One	Between	0.28	0.54	0.16	0.42	0.01	0.03
					Within	0.34	0.59	0.16	0.47	0.01	0.04
				Separate	Between	-0.13	0.05	0.10	-0.02	0.01	0.02
					Within	-0.01	0.16	0.11	0.07	0.01	0.02
		0.5	Dichotomous	Composite	Between	-0.06	0.22	0.17	0.05	0.02	0.03
					Within	-0.02	0.34	0.26	0.09	0.01	0.04

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
				One	Between	-0.04	0.18	0.16	0.07	0.01	0.03
					Within	-0.02	0.34	0.26	0.11	0.01	0.04
				Separate	Between	-0.13	0.06	0.14	-0.02	0.01	0.02
					Within	-0.08	0.26	0.25	0.02	0.01	0.03
			Polytomous	Composite	Between	0.12	0.40	0.17	0.26	0.02	0.04
					Within	0.18	0.46	0.18	0.32	0.02	0.04
				One	Between	0.13	0.40	0.16	0.28	0.02	0.03
					Within	0.18	0.46	0.17	0.34	0.02	0.04
				Separate	Between	-0.12	0.05	0.10	-0.02	0.01	0.02
					Within	-0.03	0.17	0.11	0.07	0.01	0.02
		0.7	Dichotomous	Composite	Between	-0.05	0.13	0.14	0.03	0.01	0.02
					Within	-0.04	0.28	0.25	0.07	0.01	0.03
				One	Between	-0.05	0.13	0.14	0.03	0.01	0.02
					Within	-0.04	0.29	0.25	0.08	0.01	0.03
				Separate	Between	-0.12	0.06	0.13	-0.02	0.01	0.02
					Within	-0.08	0.25	0.25	0.02	0.01	0.03
			Polytomous	Composite	Between	0.05	0.30	0.14	0.15	0.01	0.03
					Within	0.08	0.37	0.17	0.22	0.02	0.03
				One	Between	0.06	0.29	0.14	0.15	0.01	0.03
					Within	0.08	0.37	0.16	0.23	0.02	0.03
				Separate	Between	-0.10	0.06	0.10	-0.02	0.01	0.02
					Within	-0.02	0.16	0.11	0.07	0.01	0.02
		1	Dichotomous	Composite	Between	-0.10	0.06	0.13	-0.02	0.00	0.02
					Within	-0.07	0.25	0.25	0.03	0.01	0.03
				One	Between	-0.10	0.06	0.13	-0.02	0.00	0.02
					Within	-0.06	0.25	0.25	0.03	0.01	0.03
				Separate	Between	-0.12	0.06	0.13	-0.02	0.00	0.02
					Within	-0.07	0.25	0.25	0.02	0.00	0.03
			Polytomous	Composite	Between	-0.10	0.06	0.10	-0.02	0.01	0.02
					Within	-0.03	0.23	0.13	0.07	0.02	0.03
				One	Between	-0.10	0.06	0.10	-0.02	0.01	0.02
					Within	-0.03	0.23	0.13	0.07	0.02	0.03
				Separate	Between	-0.11	0.05	0.10	-0.03	0.01	0.02
					Within	-0.03	0.19	0.11	0.07	0.01	0.02
	None										
		0.2	Dichotomous	Composite	Between	-0.07	0.27	0.23	0.05	0.02	0.04
					Within	-0.07	0.32	0.23	0.05	0.02	0.04
				One	Between	-0.02	0.34	0.21	0.09	0.04	0.03
					Within	-0.03	0.34	0.21	0.09	0.04	0.03
				Separate	Between	-0.08	0.08	0.12	0.00	0.00	0.02
					Within	-0.08	0.09	0.13	0.00	0.00	0.02
			Polytomous	Composite	Between	-0.01	0.60	0.21	0.41	0.06	0.05
					Within	0.02	0.61	0.21	0.41	0.06	0.05
				One	Between	0.35	0.63	0.18	0.51	0.01	0.04
					Within	0.33	0.65	0.18	0.51	0.01	0.04
				Separate	Between	-0.06	0.08	0.10	0.01	0.00	0.02
					Within	-0.06	0.08	0.10	0.01	0.00	0.02
		0.5	Dichotomous	Composite	Between	-0.05	0.20	0.15	0.05	0.01	0.02
					Within	-0.06	0.22	0.16	0.05	0.01	0.02
				One	Between	-0.02	0.17	0.16	0.07	0.01	0.03
					Within	-0.02	0.20	0.16	0.07	0.01	0.03
				Separate	Between	-0.08	0.09	0.13	0.00	0.00	0.02
					Within	-0.09	0.09	0.13	0.00	0.00	0.02
			Polytomous	Composite	Between	0.06	0.43	0.20	0.24	0.03	0.04
					Within	0.03	0.44	0.19	0.24	0.03	0.04
				One	Between	0.12	0.42	0.18	0.28	0.02	0.04
					Within	0.13	0.44	0.17	0.28	0.02	0.04
				Separate	Between	-0.05	0.08	0.10	0.01	0.00	0.02
					Within	-0.05	0.08	0.09	0.01	0.00	0.02
		0.7	Dichotomous	Composite	Between	-0.04	0.14	0.14	0.04	0.00	0.02
					Within	-0.04	0.16	0.14	0.04	0.00	0.02
				One	Between	-0.04	0.15	0.14	0.05	0.00	0.02
					Within	-0.03	0.16	0.14	0.05	0.00	0.02
				Separate	Between	-0.09	0.09	0.13	0.00	0.00	0.02
					Within	-0.09	0.08	0.13	0.00	0.00	0.02
			Polytomous	Composite	Between	0.05	0.27	0.15	0.15	0.01	0.03
					Within	0.05	0.27	0.14	0.15	0.01	0.03
				One	Between	0.06	0.28	0.15	0.16	0.01	0.03
					Within	0.06	0.28	0.15	0.17	0.01	0.03
				Separate	Between	-0.07	0.09	0.10	0.01	0.00	0.02

Percent CPE	Level of Dependency	Correlation Between Traits	Scoring Method	Scaling Method	Comparison	Min	Max	Range	Mean	SE	SD
					Within	-0.06	0.08	0.10	0.01	0.00	0.02
		1	Dichotomous	Composite	Between	-0.09	0.08	0.13	0.00	0.00	0.02
					Within	-0.08	0.10	0.12	0.00	0.00	0.02
				One	Between	-0.09	0.08	0.13	0.00	0.00	0.02
					Within	-0.08	0.10	0.12	0.00	0.00	0.02
				Separate	Between	-0.08	0.08	0.13	0.00	0.00	0.02
					Within	-0.08	0.09	0.12	0.00	0.00	0.02
			Polytomous	Composite	Between	-0.07	0.07	0.10	0.00	0.00	0.02
					Within	-0.07	0.07	0.10	-0.01	0.00	0.02
				One	Between	-0.07	0.07	0.10	-0.01	0.00	0.02
					Within	-0.07	0.07	0.10	-0.01	0.00	0.02
				Separate	Between	-0.06	0.08	0.10	0.01	0.00	0.02
					Within	-0.05	0.08	0.10	0.01	0.00	0.02

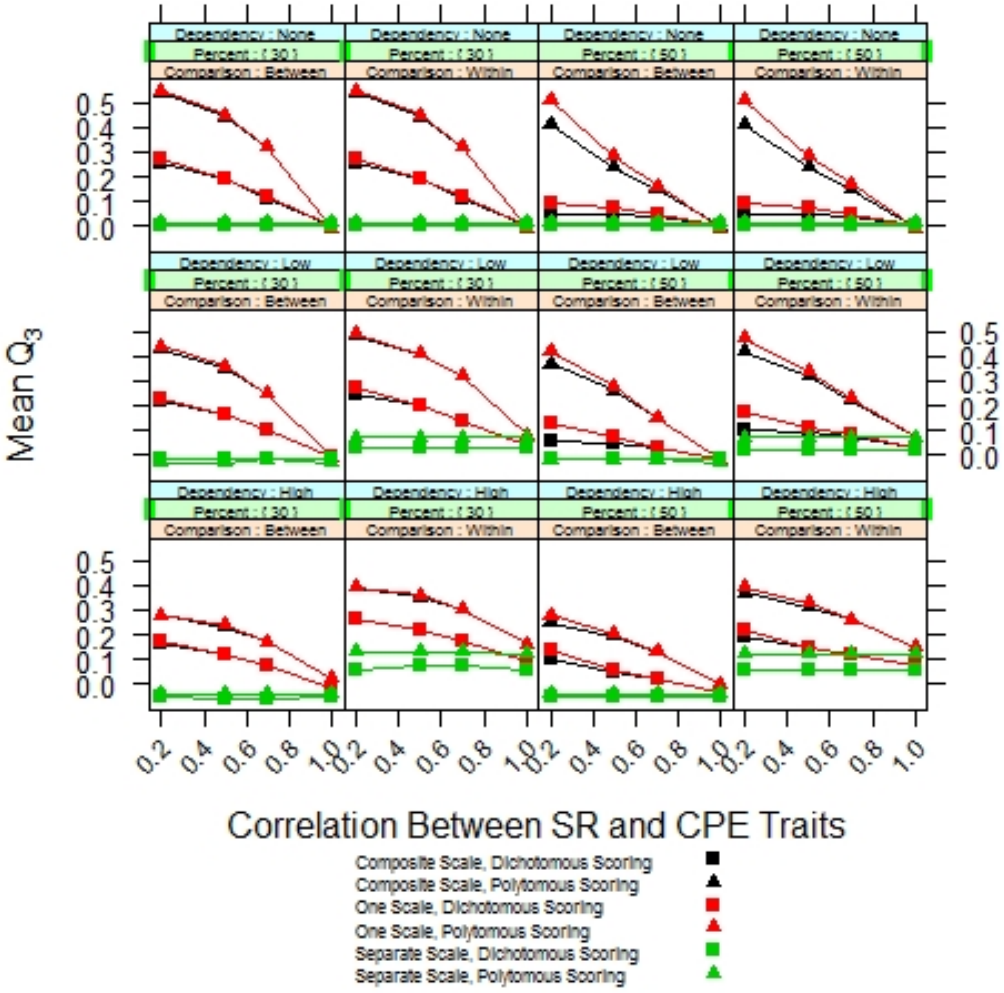

Figure 34: Average Q3 for Sample Size of 3000 and 120 Items

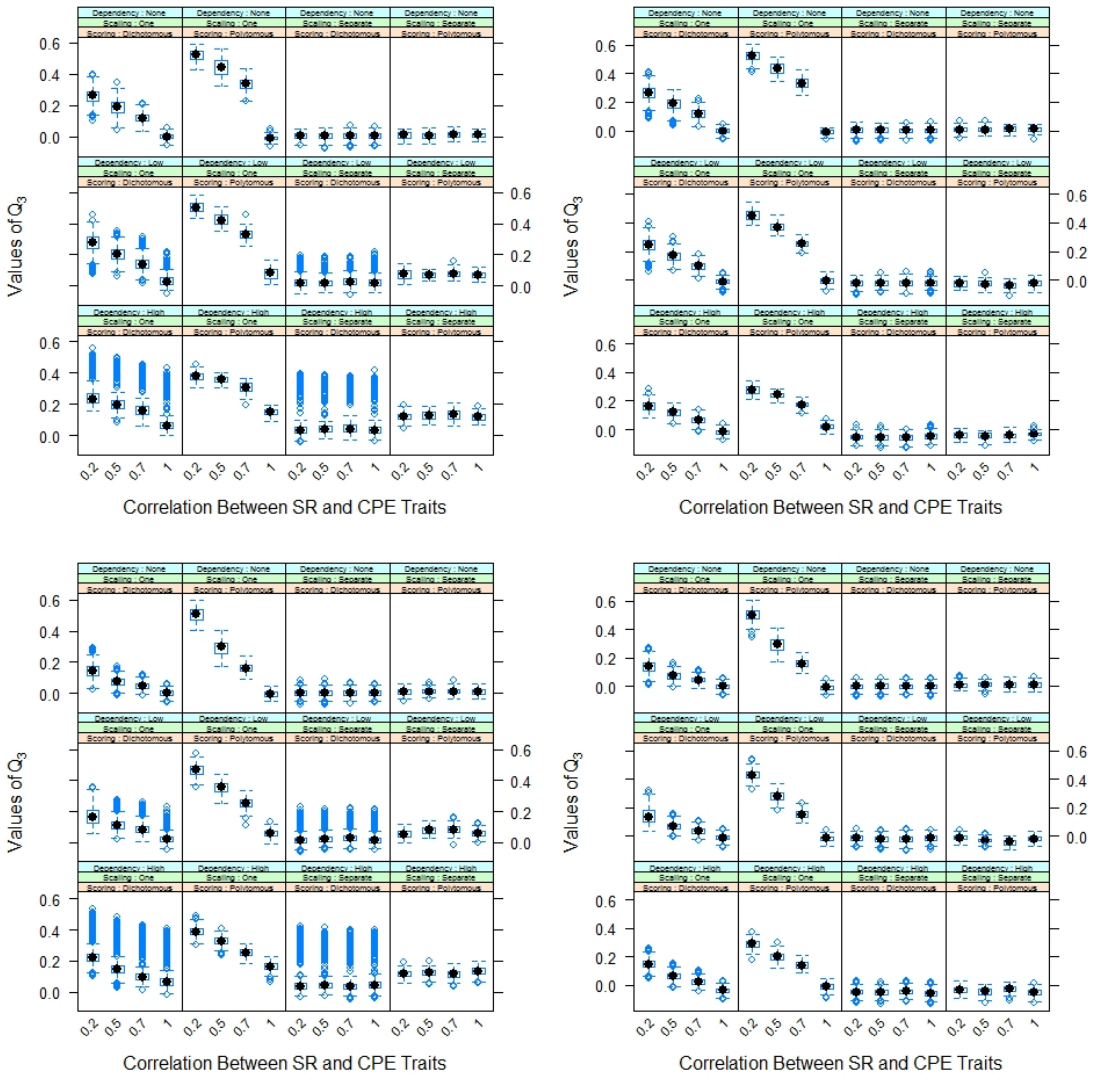

Figure 35: A-D 3000 and 120 Items for Within and Between Comparisons for 30% and 50% CPEs

Note: Clockwise from Top Left: 1) Within, 30%, 2) Between, 50%, 3) Within, 50%, 4) Between, 50%